Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2910969 C 2017/10/17

(11)(21) **2 910 969**

(12) BREVET CANADIEN / CANADIAN PATENT

(13) **C**

(86) **Date de dépôt PCT/PCT Filing Date:** 2014/06/09
(87) **Date publication PCT/PCT Publication Date:** 2014/12/18
(45) **Date de délivrance/Issue Date:** 2017/10/17
(85) **Entrée phase nationale/National Entry:** 2015/10/29
(86) **N° demande PCT/PCT Application No.:** US 2014/041505
(87) **N° publication PCT/PCT Publication No.:** 2014/200899
(30) **Priorités/Priorities:** 2013/06/11 (US61/833,836); 2013/07/29 (US61/859,702); 2014/06/06 (US14/298,555)

(51) **Cl.Int./Int.Cl.** ***H04N 19/70*** (2014.01)

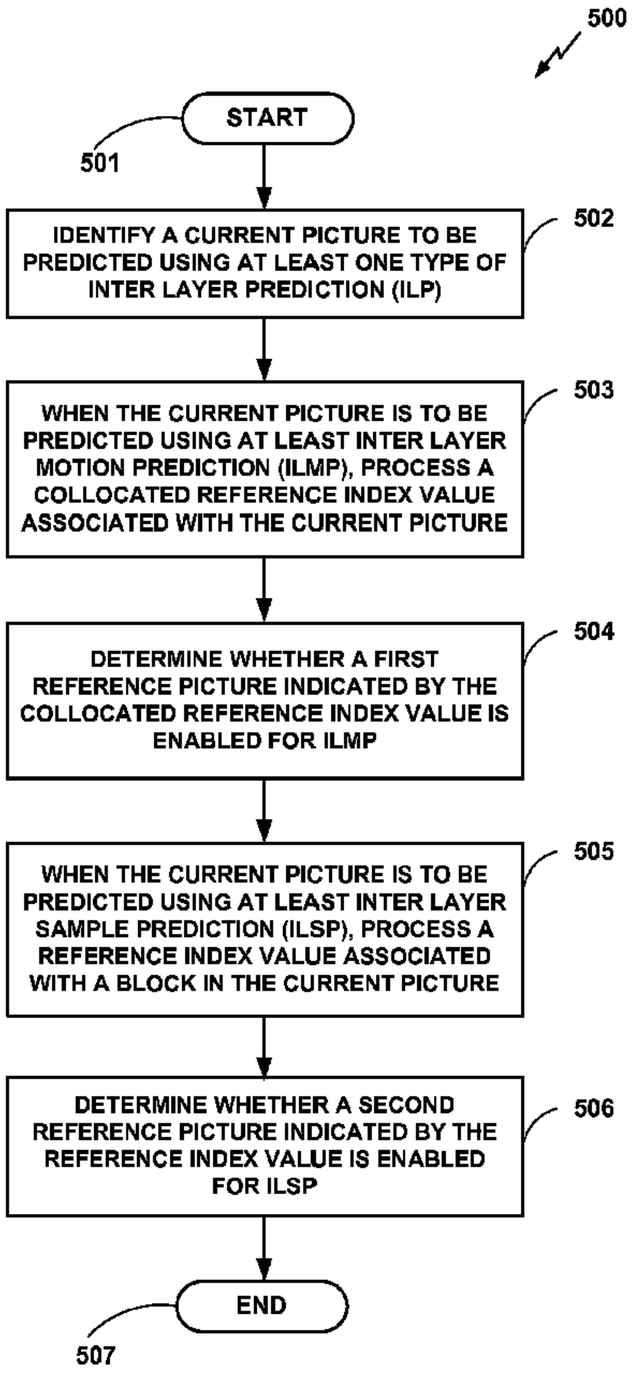
(72) **Inventeurs/Inventors:**
SEREGIN, VADIM, US;
WANG, YE-KUI, US

(73) **Propriétaire/Owner:**
QUALCOMM INCORPORATED, US

(74) **Agent:** SMART & BIGGAR

(54) **Titre : CONTRAINTES SUR LES FLUX BINAIRES DE TRAITEMENT LIEES AUX TYPES DE PREDICTION INTER-COUCHES EN CODAGE VIDEO MULTICOUCHE**

(54) **Title: PROCESSING BITSTREAM CONSTRAINTS RELATING TO INTER-LAYER PREDICTION TYPES IN MULTI-LAYER VIDEO CODING**



(57) **Abrégé/Abstract:**
An apparatus for coding video information may include computing hardware configured to: when a current picture is to be predicted using at least inter layer motion prediction (ILMP): process a collocated reference index value associated with the current

CA 2910969 C 2017/10/17

(11)(21) **2 910 969**

(13) **C**

(57) **Abrégé(suite)/Abstract(continued):**
picture, wherein the collocated reference index value indicates a first reference picture that is used in predicting the current picture using inter layer prediction (ILP); and determine whether the first reference picture indicated by the collocated reference index value is enabled for ILMP; when the current picture is to be predicted using at least inter layer sample prediction (ILSP): process a reference index value associated with a block in the current picture, wherein the reference index value indicates a second reference picture that is used in predicting the block in the current picture using ILP; and determine whether the second reference picture indicated by the reference index value is enabled for ILSP.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 December 2014 (18.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/200899 A1

(51) International Patent Classification:
H04N 19/70 (2014.01)

(21) International Application Number:
PCT/US2014/041505

(22) International Filing Date:
9 June 2014 (09.06.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/833,836	11 June 2013 (11.06.2013)	US
61/859,702	29 July 2013 (29.07.2013)	US
14/298,555	6 June 2014 (06.06.2014)	US

(71) Applicant: **QUALCOMM INCORPORATED** [US/US]; ATTN: International IP Administration, 5775 Morehouse Drive, San Diego, California 92121-1714 (US).

(72) Inventors: **SEREGIN, Vadim**; 5775 Morehouse Drive, San Diego, California 92121-1714 (US). **WANG, Ye-Kui**; 5775 Morehouse Drive, San Diego, California 92121-1714 (US).

(74) Agent: **ABUMERI, Mark M.**; Knobbe Martens Olson & Bear LLP, 2040 Main Street, Fourteenth Floor, Irvine, California 92614 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

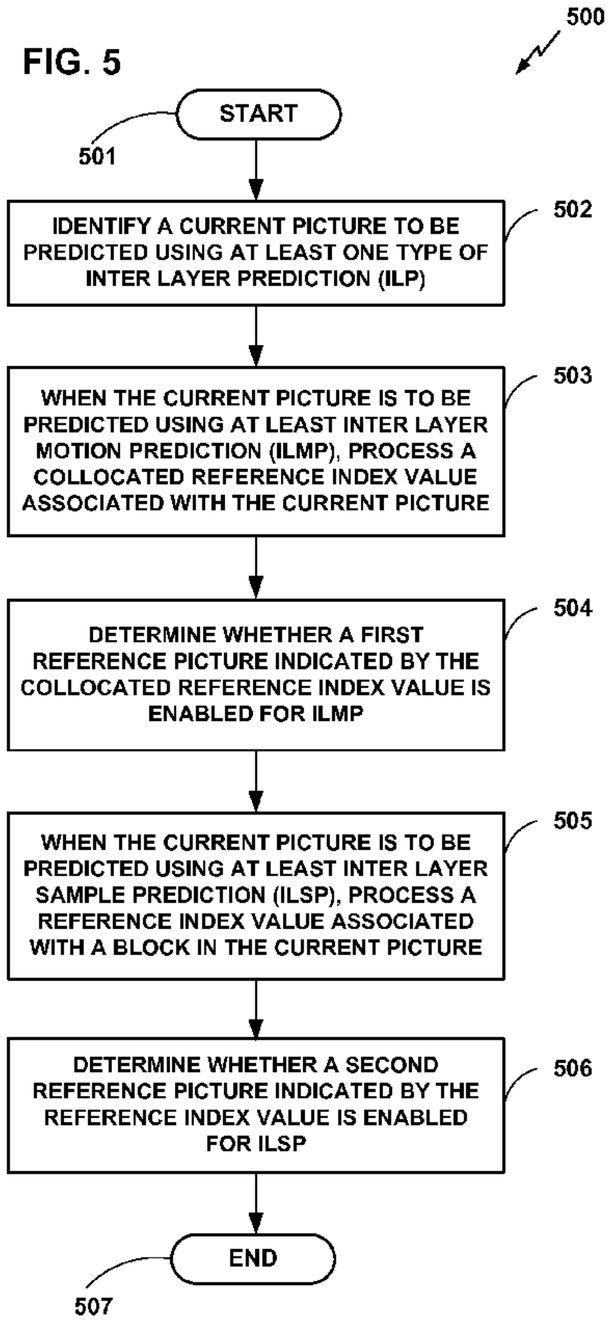
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,

[Continued on next page]

(54) Title: PROCESSING BITSTREAM CONSTRAINTS RELATING TO INTER-LAYER PREDICTION TYPES IN MULTI-LAYER VIDEO CODING

(57) Abstract: An apparatus for coding video information may include computing hardware configured to: when a current picture is to be predicted using at least inter layer motion prediction (ILMP): process a collocated reference index value associated with the current picture, wherein the collocated reference index value indicates a first reference picture that is used in predicting the current picture using inter layer prediction (ILP); and determine whether the first reference picture indicated by the collocated reference index value is enabled for ILMP; when the current picture is to be predicted using at least inter layer sample prediction (ILSP): process a reference index value associated with a block in the current picture, wherein the reference index value indicates a second reference picture that is used in predicting the block in the current picture using ILP; and determine whether the second reference picture indicated by the reference index value is enabled for ILSP.

FIG. 5



WO 2014/200899 A1

WO 2014/200899 A1

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

PROCESSING BITSTREAM CONSTRAINTS RELATING TO INTER-LAYER PREDICTION TYPES IN MULTI-LAYER VIDEO CODING

BACKGROUND

Field

[0001] This disclosure is related to the field of video coding and compression. In particular, it is related to scalable video coding (SVC), including SVC for Advanced Video Coding (AVC), as well as SVC for High Efficiency Video Coding (HEVC), which is also referred to as Scalable HEVC (SHVC). It is also related to 3D video coding, such as the multiview extension of HEVC, referred to as MV-HEVC. Various embodiments relate to systems and methods for independent control of inter-layer motion prediction reference resampling and inter-layer sample prediction reference resampling and for processing bitstream constraints relating to inter-layer prediction types.

Description of the Related Art

[0002] Digital video capabilities can be incorporated into a wide range of devices, including digital televisions, digital direct broadcast systems, wireless broadcast systems, personal digital assistants (PDAs), laptop or desktop computers, tablet computers, e-book readers, digital cameras, digital recording devices, digital media players, video gaming devices, video game consoles, cellular or satellite radio telephones, so-called "smart phones," video teleconferencing devices, video streaming devices, and the like. Digital video devices implement video coding techniques, such as those described in the standards defined by MPEG-2, MPEG-4, ITU-T H.263, ITU-T H.264/MPEG-4, Part 10, Advanced Video Coding (AVC), the High Efficiency Video Coding (HEVC) standard presently under development, and extensions of such standards. The video devices may transmit, receive, encode, decode, and/or store digital video information more efficiently by implementing such video coding techniques.

[0003] Video coding techniques include spatial (intra-picture) prediction and/or temporal (inter-picture) prediction to reduce or remove redundancy inherent in video sequences. For block-based video coding, a video slice (e.g., a video frame or a portion of a video frame) may be partitioned into video blocks, which may also be referred to as treeblocks, coding units (CUs) and/or coding nodes. Video blocks in an

intra-coded (I) slice of a picture are encoded using spatial prediction with respect to reference samples in neighboring blocks in the same picture. Video blocks in an inter-coded (P or B) slice of a picture may use spatial prediction with respect to reference samples in neighboring blocks in the same picture or temporal prediction with respect to reference samples in other reference pictures. Pictures may be referred to as frames, and reference pictures may be referred to a reference frames.

[0004] Spatial or temporal prediction results in a predictive block for a block to be coded. Residual data represents pixel differences between the original block to be coded and the predictive block. An inter-coded block is encoded according to a motion vector that points to a block of reference samples forming the predictive block, and the residual data indicating the difference between the coded block and the predictive block. An intra-coded block is encoded according to an intra-coding mode and the residual data. For further compression, the residual data may be transformed from the pixel domain to a transform domain, resulting in residual transform coefficients, which then may be quantized. The quantized transform coefficients, initially arranged in a two-dimensional array, may be scanned in order to produce a one-dimensional vector of transform coefficients, and entropy coding may be applied to achieve even more compression.

SUMMARY

[0005] In general, this disclosure describes techniques related to scalable video coding (SVC). Various techniques described below provide describe methods and devices for independent control of inter-layer motion prediction reference resampling and inter-layer sample prediction reference resampling. Various techniques described below provide describe methods and devices for processing bitstream constraints relating to inter-layer prediction types.

[0006] An apparatus for coding video information according to certain aspects includes a memory and computing hardware. The memory unit is configured to store video information. The computing hardware is configured to: identify a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising one or more of inter layer motion prediction (ILMP) or inter layer sample prediction (ILSP); and control: (1) a number of pictures that may be resampled and used to predict the current picture using ILMP and (2) a number of pictures that

may be resampled and used to predict the current picture using ILSP, wherein the computing hardware is configured to control the number of pictures that may be resampled and used to predict the current picture using ILMP independent of the number of pictures that may be resampled and used to predict the current picture using ILSP.

[0007] An apparatus for coding video information according to certain aspects includes a memory and computing hardware. The memory unit is configured to store video information. The computing hardware is configured to: identify a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both; when the current picture is to be predicted using at least ILMP: process a collocated reference index value associated with the current picture, wherein the collocated reference index value indicates a first reference picture that is used in predicting the current picture using ILMP; and determine whether the first reference picture indicated by the collocated reference index value is enabled for ILMP; and when the current picture is to be predicted using at least ILSP: process a reference index value associated with a block in the current picture, wherein the reference index value indicates a second reference picture that is used in predicting the block in the current picture using ILSP; and determine whether the second reference picture indicated by the reference index value is enabled for ILSP.

[0007a] According to one aspect of the present invention, there is provided an apparatus configured to code video information, the apparatus comprising: a memory configured to store video data; and computing hardware operationally coupled to the memory and configured to: identify a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both; in response to a determination that the current picture is to be predicted using at least ILMP: process a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP; determine whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the

current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; and in response to a determination that the current picture is to be predicted using at least ILSP: process a reference index value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP; and determine whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

[0007b] According to another aspect of the present invention, there is provided A method of coding video information, the method comprising: identifying a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both; and performing at least one of: (i) in response to a determination that the current picture is to be predicted using at least ILMP: processing a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP; and determining whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; or (ii) in response to a determination that the current picture is to be predicted using at least ILSP: processing a reference value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP; and
determining whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

[0007c] According to still another aspect of the present invention, there is provided a non-transitory computer readable medium comprising instructions that when executed on a processor comprising computer hardware cause the processor to: identify a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both; and perform at least one of: (i) in response to a determination that the current picture is to be predicted using at least ILMP: process a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP; and determine whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; or (ii) in response to a determination that the current picture is to be predicted using at least ILSP: process a reference index value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP; and determine whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

[0007d] According to a further aspect of the present invention, there is provided an apparatus configured to code video information, the apparatus comprising: means for storing video data; means for identifying a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both; means for, in response to a determination that the current picture is to be predicted using at least ILMP, processing a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP, the means

for processing the collocated reference index value further configured to determine whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; and means for, in response to a determination that the current picture is to be predicted using at least ILSP, processing a reference index value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP, the means for processing the reference index value further configured to determine whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

[0008] The details of one or more examples are set forth in the accompanying drawings and the description below, which are not intended to limit the full scope of the inventive concepts described herein. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Throughout the drawings, reference numbers may be re-used to indicate correspondence between referenced elements. The drawings are provided to illustrate example embodiments described herein and are not intended to limit the scope of the disclosure.

[0010] FIG. 1 is a block diagram illustrating an example video encoding and decoding system that may utilize techniques in accordance with aspects described in this disclosure.

[0011] FIG. 2A is a block diagram illustrating an example of a video encoder that may implement techniques in accordance with aspects described in this disclosure.

[0012] FIG. 2B is a block diagram illustrating an example of a video encoder that may implement techniques in accordance with aspects described in this disclosure.

[0013] FIG. 3A is a block diagram illustrating an example of a video decoder that may implement techniques in accordance with aspects described in this disclosure.

[0014] FIG. 3B is a block diagram illustrating an example of a video decoder that may implement techniques in accordance with aspects described in this disclosure.

[0015] FIG. 4 is a flowchart illustrating an example method for independent control of inter-layer motion prediction reference resampling and inter-layer sample prediction reference resampling, according to aspects of this disclosure.

[0016] FIG. 5 is a flowchart illustrating an example method for processing bitstream constraints relating to inter-layer prediction types.

DETAILED DESCRIPTION

[0017] The techniques described in this disclosure generally relate to scalable video coding (SHVC, SVC) and multiview/3D video coding (e.g., multiview coding plus depth, MVC+D). For example, the techniques may be related to, and used with or within, a High Efficiency Video Coding (HEVC) scalable video coding (SVC, sometimes referred to as SHVC) extension. In an SHVC, SVC extension, there could be multiple layers of video information. The layer at the lowest level of the video information may serve as a base layer (BL) or reference layer (RL), and the layer at the very top (or the highest layer) of the video information may serve as an enhanced layer (EL). The "enhanced layer" is sometimes referred to as an "enhancement layer," and these terms may be used interchangeably. The base layer is sometimes referred to as a "reference layer," and these terms may also be used interchangeably. All layers in between the base layer and the top layer may serve as additional ELs and/or reference layers. For example, a given layer may be an EL for a layer below (e.g., that precedes) the given layer, such as the base layer or any intervening enhancement layer. Further, the given layer may also serve as a RL for one or more the enhancement layer(s) above (e.g., subsequent to) the given layer. Any layer in between the base layer (e.g., the lowest layer having, for example, a layer identification (ID) set or equal to "1") and the top layer (or the highest layer) may be used as a reference for inter-layer prediction by a

layer higher to the given layer and may use a layer lower to the given layer as a reference for inter-layer prediction. For example, the given layer can be determined using a layer lower to the given layer as a reference for inter-layer prediction.

[0018] For simplicity, examples are presented in terms of just two layers: a BL and an EL; however, it should be well understood that the ideas and embodiments described below are applicable to cases with multiple layers, as well. In addition, for ease of explanation, the terms "frames" or "blocks" are often used. However, these terms are not meant to be limiting. For example, the techniques described below can be used with any of a variety of video units, including but not limited to pixels, blocks (e.g., CU, PU, TU, macroblocks, etc.), slices, frames, picture, etc.

Video Coding

[0019] Video coding standards include ITU-T H.261, ISO/IEC MPEG-1 Visual, ITU-T H.262 or ISO/IEC MPEG-2 Visual, ITU-T H.263, ISO/IEC MPEG-4 Visual and ITU-T H.264 (also known as ISO/IEC MPEG-4 AVC), including its Scalable Video Coding (SVC) and Multi-view Video Coding (MVC) and Multi-view Coding plus Depth (MVC+D) extensions. The latest HEVC draft specification, and referred to as HEVC WD10 hereinafter, is available from http://phenix.int-evry.fr/jct/doc_end_user/documents/12_Geneva/wg11/JCTVC-L1003-v34.zip. The multiview extension to HEVC, namely MV-HEVC, is also being developed by the JCT-3V. A recent Working Draft (WD) of MV-HEVC WD3 hereinafter, is available from http://phenix.it-sudparis.eu/jct2/doc_end_user/documents/3_Geneva/wg11/JCT3V-C1004-v4.zip. The scalable extension to HEVC, named SHVC, is also being developed by the JCT-VC. A recent Working Draft (WD) of SHVC and referred to as SHVC WD2 hereinafter, is available from http://phenix.int-evry.fr/jct/doc_end_user/documents/13_Incheon/wg11/JCTVC-M1008-v1.zip.

[0020] In SVC and SHVC, video information may be provided as multiple layers. The layer at the very bottom level can just serve as a base layer (BL) and the layer at the very top level can serve as an enhancement layer (EL). All the layers between the top and bottom layers may serve as both enhancement layers and reference layers. For example, a layer in the middle can be an EL for the layers below it, and at the same time as a RL for the layers above it. For simplicity of description, we can assume that there are two layers, a BL and an EL, in illustrating the techniques

described below. However, all the techniques described herein are applicable to cases with multiple (more than two) layers, as well.

[0021] Scalable video coding (SVC) may be used to provide quality (also referred to as signal-to-noise (SNR)) scalability, spatial scalability and/or temporal scalability. For example, in one embodiment, a reference layer (e.g., a base layer) includes video information sufficient to display a video at a first quality level and the enhancement layer includes additional video information relative to the reference layer such that the reference layer and the enhancement layer together include video information sufficient to display the video at a second quality level higher than the first level (e.g., less noise, greater resolution, better frame rate, etc.). An enhanced layer may have different spatial resolution than a base layer. For example, the spatial aspect ratio between EL and BL can be 1.0, 1.5, 2.0 or other different ratios. In other words, the spatial aspect of the EL may equal 1.0, 1.5, or 2.0 times the spatial aspect of the BL. In some examples, the scaling factor of the EL may be greater than the BL. For example, a size of pictures in the EL may be greater than a size of pictures in the BL. In this way, it may be possible, although not a limitation, that the spatial resolution of the EL is larger than the spatial resolution of the BL.

[0022] In SVC, which refers to the SVC extension for H.264 or the SHVC extension for H.265 (as discussed above), prediction of a current block may be performed using the different layers that are provided for SVC. Such prediction may be referred to as inter-layer prediction. Inter-layer prediction methods may be utilized in SVC in order to reduce inter-layer redundancy. Some examples of inter-layer prediction may include inter-layer intra prediction, inter-layer motion prediction, and inter-layer residual prediction. Inter-layer intra prediction uses the reconstruction of co-located blocks in the base layer to predict the current block in the enhancement layer. Inter-layer motion prediction uses motion information (including motion vectors) of the base layer to predict motion in the enhancement layer. Inter-layer residual prediction uses the residue of the base layer to predict the residue of the enhancement layer.

Overview

[0023] In SHVC, an inter-layer reference picture (ILRP) used in inter-layer prediction (ILP) may be used for inter-layer motion prediction (ILMP), inter-layer sample prediction (ILSP), or both. The type of ILP an ILRP is used for can be referred

to as inter-layer prediction type (e.g., ILMP, ILSP, or both). For a reference picture used for ILSP only, if the reference layer picture has a different picture size from the current picture, the reference layer picture should be sample-resampled to generate the ILRP, but not motion-resampled since motion information is not used. For a reference picture used for ILMP only, if the reference layer picture has a different picture size from the current picture, the reference layer picture should be motion-resampled to generate ILRP, but not sample-resampled since samples from the reference layer picture are not used. For a reference picture used for both ILSP and ILMP, if the reference picture has a different size from the current picture, the reference layer picture should be sample-resampled and motion-resampled.

[0024] In the early versions of SHVC Working Draft (WD), if a reference layer picture has a different size from the current picture, the resampling process is invoked to derive the inter-layer reference picture without checking the inter-layer prediction type (e.g., ILMP, ILSP, or both) of the reference layer picture. This can lead to sample-resampling an ILRP used only for ILMP even though the samples from the ILRP are not needed. Moreover, in some SHVC profiles, the number of inter-layer reference pictures that can be resampled for decoding any particular picture can be limited to a certain number (e.g., 1). However, the two resampling processes (e.g., sample-resampling and motion-resampling) were not separately considered in counting the number of resampled pictures. Accordingly, if the sample-resampling process is invoked for a picture used for inter-layer motion prediction only, then the sample-resampling process can no longer be invoked for another picture for inter-layer sample prediction when decoding the particular picture. Therefore, it would be advantageous to not sample-resample an ILRP used only for ILMP, and also to not count the sample-resampling of an ILRP used only for ILMP toward the limit on number of ILRP resampled for a particular picture. In another example, if the motion-resampling process is invoked for a picture used for ILSP only, then the motion-resampling process can no longer be invoked for another picture for ILMP when decoding the particular picture. It would also be advantageous to not motion-resample an ILRP used only for ILSP, and also to not count the motion-resampling of an ILRP used only for ILSP toward the limit on number of ILRP resampled for a particular picture. To facilitate discussion, the limit on number of ILRP resampled for a particular picture may also be referred to as “resampled picture count.”

[0025] In order to address these and other challenges, the techniques can avoid invoking the resampling process for inter-layer reference pictures used for inter-layer motion prediction only. The techniques can also not count inter-layer reference pictures used for only inter-layer motion prediction towards the resampled picture count even when the ILRPs have different picture size from the current picture.

[0026] In certain embodiments, the techniques can count inter-layer reference pictures used for inter-layer motion prediction separately from the inter-layer reference pictures used for inter-layer sample prediction with respect to the constraint on the number of the resampled pictures. For example, the techniques can have one resampled picture count for ILRPs for ILMP, and another resampled picture count for ILRPs for ILSP.

[0027] In addition, the techniques can also provide and/or process bitstream constraints relating to inter-layer prediction types. For example, the techniques can provide and/or process a bitstream constraint that the collocated reference index (e.g., collocated_ref_idx) can only refer to an ILRP used for at least ILMP. The techniques can also provide and/or process a bitstream constraint that reference index (e.g., ref_idx) can only refer to an ILRP used for at least ILSP. The bitstream constraints can be implemented using one or more flags.

[0028] Various aspects of the novel systems, apparatuses, and methods are described more fully hereinafter with reference to the accompanying drawings. This disclosure may, however, be embodied in many different forms and should not be construed as limited to any specific structure or function presented throughout this disclosure. Rather, these aspects are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the disclosure to those skilled in the art. Based on the teachings herein one skilled in the art should appreciate that the scope of the disclosure is intended to cover any aspect of the novel systems, apparatuses, and methods disclosed herein, whether implemented independently of, or combined with, any other aspect of the invention. For example, an apparatus may be implemented or a method may be practiced using any number of the aspects set forth herein. In addition, the scope of the invention is intended to cover such an apparatus or method which is practiced using other structure, functionality, or structure and functionality in addition to or other than the various aspects of the invention set forth herein. It should be

understood that any aspect disclosed herein may be embodied by one or more elements of a claim.

[0029] Although particular aspects are described herein, many variations and permutations of these aspects fall within the scope of the disclosure. Although some benefits and advantages of the preferred aspects are mentioned, the scope of the disclosure is not intended to be limited to particular benefits, uses, or objectives. Rather, aspects of the disclosure are intended to be broadly applicable to different wireless technologies, system configurations, networks, and transmission protocols, some of which are illustrated by way of example in the figures and in the following description of the preferred aspects. The detailed description and drawings are merely illustrative of the disclosure rather than limiting, the scope of the disclosure being defined by the appended claims and equivalents thereof.

Video Coding System

[0030] FIG. 1 is a block diagram that illustrates an example video coding system 10 that may utilize techniques in accordance with aspects described in this disclosure. As used described herein, the term "video coder" refers generically to both video encoders and video decoders. In this disclosure, the terms "video coding" or "coding" may refer generically to video encoding and video decoding.

[0031] As shown in FIG. 1, video coding system 10 includes a source device 12 and a destination device 14. Source device 12 generates encoded video data. Destination device 14 may decode the encoded video data generated by source device 12. Source device 12 can provide the video data to the destination device 14 via a communication channel 16, which may include a computer-readable storage medium or other communication channel. Source device 12 and destination device 14 may include a wide range of devices, including desktop computers, notebook (e.g., laptop) computers, tablet computers, set-top boxes, telephone handsets, such as so-called "smart" phones, so-called "smart" pads, televisions, cameras, display devices, digital media players, video gaming consoles, in-car computers, video streaming devices, or the like. Source device 12 and destination device 14 may be equipped for wireless communication.

[0032] Destination device 14 may receive the encoded video data to be decoded via communication channel 16. Communication channel 16 may comprise a

type of medium or device capable of moving the encoded video data from source device 12 to destination device 14. For example, communication channel 16 may comprise a communication medium to enable source device 12 to transmit encoded video data directly to destination device 14 in real-time. The encoded video data may be modulated according to a communication standard, such as a wireless communication protocol, and transmitted to destination device 14. The communication medium may comprise a wireless or wired communication medium, such as a radio frequency (RF) spectrum or one or more physical transmission lines. The communication medium may form part of a packet-based network, such as a local area network, a wide-area network, or a global network, such as the Internet. The communication medium may include routers, switches, base stations, or other equipment that may be useful to facilitate communication from source device 12 to destination device 14.

[0033] In some embodiments, encoded data may be output from output interface 22 to a storage device. In such examples, channel 16 may correspond to a storage device or computer-readable storage medium that stores the encoded video data generated by source device 12. For example, destination device 14 may access the computer-readable storage medium via disk access or card access. Similarly, encoded data may be accessed from the computer-readable storage medium by input interface 28. The computer-readable storage medium may include any of a variety of distributed or locally accessed data storage media such as a hard drive, Blu-ray discs, DVDs, CD-ROMs, flash memory, volatile or non-volatile memory, or other digital storage media for storing video data. The computer-readable storage medium may correspond to a file server or another intermediate storage device that may store the encoded video generated by source device 12. Destination device 14 may access stored video data from the computer-readable storage medium via streaming or download. The file server may be a type of server capable of storing encoded video data and transmitting that encoded video data to the destination device 14. Example file servers include a web server (e.g., for a website), an FTP server, network attached storage (NAS) devices, or a local disk drive. Destination device 14 may access the encoded video data through a standard data connection, including an Internet connection. This may include a wireless channel (e.g., a Wi-Fi connection), a wired connection (e.g., DSL, cable modem, etc.), or a combination of both that is suitable for accessing encoded video data stored on a file server. The transmission of encoded video data from the computer-readable storage

medium may be a streaming transmission, a download transmission, or a combination of both.

[0034] The techniques of this disclosure can apply applications or settings in addition to wireless applications or settings. The techniques may be applied to video coding in support of a of a variety of multimedia applications, such as over-the-air television broadcasts, cable television transmissions, satellite television transmissions, Internet streaming video transmissions, such as dynamic adaptive streaming over HTTP (DASH), digital video that is encoded onto a data storage medium, decoding of digital video stored on a data storage medium, or other applications. In some embodiments, system 10 may be configured to support one-way or two-way video transmission to support applications such as video streaming, video playback, video broadcasting, and/or video telephony.

[0035] In FIG. 1, source device 12 includes video source 18, video encoder 20, and output interface 22. Destination device 14 includes input interface 28, video decoder 30, and display device 32. Video encoder 20 of source device 12 may be configured to apply the techniques for coding a bitstream including video data conforming to multiple standards or standard extensions. In other embodiments, a source device and a destination device may include other components or arrangements. For example, source device 12 may receive video data from an external video source 18, such as an external camera. Likewise, destination device 14 may interface with an external display device, rather than including an integrated display device.

[0036] Video source 18 of source device 12 may include a video capture device, such as a video camera, a video archive containing previously captured video, and/or a video feed interface to receive video from a video content provider. Video source 18 may generate computer graphics-based data as the source video, or a combination of live video, archived video, and computer-generated video. In some embodiments, if video source 18 is a video camera, source device 12 and destination device 14 may form so-called camera phones or video phones. The captured, pre-captured, or computer-generated video may be encoded by video encoder 20. The encoded video information may be output by output interface 22 to a communication channel 16, which may include a computer-readable storage medium, as discussed above.

[0037] Computer-readable storage medium may include transient media, such as a wireless broadcast or wired network transmission, or storage media (e.g., non-transitory storage media), such as a hard disk, flash drive, compact disc, digital video disc, Blu-ray disc, or other computer-readable media. A network server (not shown) may receive encoded video data from source device 12 and provide the encoded video data to destination device 14 (e.g., via network transmission). A computing device of a medium production facility, such as a disc stamping facility, may receive encoded video data from source device 12 and produce a disc containing the encoded video data. Therefore, communication channel 16 may be understood to include one or more computer-readable storage media of various forms.

[0038] Input interface 28 of destination device 14 can receive information from communication channel 16. The information of communication channel 16 may include syntax information defined by video encoder 20, which can be used by video decoder 30, that includes syntax elements that describe characteristics and/or processing of blocks and other coded units, e.g., GOPs. Display device 32 displays the decoded video data to a user, and may include any of a variety of display devices such as a cathode ray tube (CRT), a liquid crystal display (LCD), a plasma display, an organic light emitting diode (OLED) display, or another type of display device.

[0039] Video encoder 20 and video decoder 30 may operate according to a video coding standard, such as the High Efficiency Video Coding (HEVC) standard presently under development, and may conform to the HEVC Test Model (HM). Alternatively, video encoder 20 and video decoder 30 may operate according to other proprietary or industry standards, such as the ITU-T H.264 standard, alternatively referred to as MPEG-4, Part 10, Advanced Video Coding (AVC), or extensions of such standards. The techniques of this disclosure, however, are not limited to any particular coding standard. Other examples of video coding standards include MPEG-2 and ITU-T H.263. Although not shown in FIG. 1, in some aspects, video encoder 20 and video decoder 30 may each be integrated with an audio encoder and decoder, and may include appropriate MUX-DEMUX units, or other hardware and software, to handle encoding of both audio and video in a common data stream or separate data streams. If applicable, MUX-DEMUX units may conform to the ITU H.223 multiplexer protocol, or other protocols such as the user datagram protocol (UDP).

[0040] FIG. 1 is merely an example and the techniques of this disclosure may apply to video coding settings (e.g., video encoding or video decoding) that do not necessarily include any data communication between the encoding and decoding devices. In other examples, data can be retrieved from a local memory, streamed over a network, or the like. An encoding device may encode and store data to memory, and/or a decoding device may retrieve and decode data from memory. In many examples, the encoding and decoding is performed by devices that do not communicate with one another, but simply encode data to memory and/or retrieve and decode data from memory.

[0041] Video encoder 20 and video decoder 30 each may be implemented as any of a variety of suitable encoder circuitry, such as one or more microprocessors, digital signal processors (DSPs), application specific integrated circuits (ASICs), field programmable gate arrays (FPGAs), discrete logic, software, hardware, firmware or any combinations thereof. When the techniques are implemented partially in software, a device may store instructions for the software in a non-transitory computer-readable medium and execute the instructions in hardware using one or more processors to perform the techniques of this disclosure. Each of video encoder 20 and video decoder 30 may be included in one or more encoders or decoders, either of which may be integrated as part of a combined encoder/decoder (CODEC) in a respective device. A device including video encoder 20 and/or video decoder 30 may comprise an integrated circuit, a microprocessor, and/or a wireless communication device, such as a cellular telephone.

[0042] The JCT-VC is working on development of the HEVC standard and its extension and Version 1 has been finalized. The HEVC standardization efforts are based on an evolving model of a video coding device referred to as the HEVC Test Model (HM). The HM presumes several additional capabilities of video coding devices relative to existing devices according to, e.g., ITU-T H.264/AVC. For example, whereas H.264 provides nine intra-prediction encoding modes, the HM may provide as many as thirty-three intra-prediction encoding modes.

[0043] In general, the working model of the HM describes that a video frame or picture may be divided into a sequence of treeblocks or largest coding units (LCU) that include both luma and chroma samples. Syntax data within a bitstream may define a size for the LCU, which is a largest coding unit in terms of the number of

pixels. A slice includes a number of consecutive treeblocks in coding order. A video frame or picture may be partitioned into one or more slices. Each treeblock may be split into coding units (CUs) according to a quadtree. In general, a quadtree data structure includes one node per CU, with a root node corresponding to the treeblock. If a CU is split into four sub-CUs, the node corresponding to the CU includes four leaf nodes, each of which corresponds to one of the sub-CUs.

[0044] Each node of the quadtree data structure may provide syntax data for the corresponding CU. For example, a node in the quadtree may include a split flag, indicating whether the CU corresponding to the node is split into sub-CUs. Syntax elements for a CU may be defined recursively, and may depend on whether the CU is split into sub-CUs. If a CU is not split further, it is referred as a leaf-CU. In this disclosure, four sub-CUs of a leaf-CU will also be referred to as leaf-CUs even if there is no explicit splitting of the original leaf-CU. For example, if a CU at 16x16 size is not split further, the four 8x8 sub-CUs will also be referred to as leaf-CUs although the 16x16 CU was never split.

[0045] A CU has a similar purpose as a macroblock of the H.264 standard, except that a CU does not have a size distinction. For example, a treeblock may be split into four child nodes (also referred to as sub-CUs), and each child node may in turn be a parent node and be split into another four child nodes. A final, unsplit child node, referred to as a leaf node of the quadtree, comprises a coding node, also referred to as a leaf-CU. Syntax data associated with a coded bitstream may define a maximum number of times a treeblock may be split, referred to as a maximum CU depth, and may also define a minimum size of the coding nodes. Accordingly, a bitstream may also define a smallest coding unit (SCU). This disclosure uses the term "block" to refer to any of a CU, PU, or TU, in the context of HEVC, or similar data structures in the context of other standards (e.g., macroblocks and sub-blocks thereof in H.264/AVC).

[0046] A CU includes a coding node and prediction units (PUs) and transform units (TUs) associated with the coding node. A size of the CU corresponds to a size of the coding node and must be square in shape. The size of the CU may range from 8x8 pixels up to the size of the treeblock with a maximum of 64x64 pixels or greater. Each CU may contain one or more PUs and one or more TUs. Syntax data associated with a CU may describe, for example, partitioning of the CU into one or more PUs. Partitioning modes may differ between whether the CU is skip or direct

mode encoded, intra-prediction mode encoded, or inter-prediction mode encoded. PUs may be partitioned to be non-square in shape. Syntax data associated with a CU may also describe, for example, partitioning of the CU into one or more TUs according to a quadtree. A TU can be square or non-square (e.g., rectangular) in shape.

[0047] The HEVC standard allows for transformations according to TUs, which may be different for different CUs. The TUs are typically sized based on the size of PUs within a given CU defined for a partitioned LCU, although this may not always be the case. The TUs are typically the same size or smaller than the PUs. In some examples, residual samples corresponding to a CU may be subdivided into smaller units using a quadtree structure known as "residual quad tree" (RQT). The leaf nodes of the RQT may be referred to as transform units (TUs). Pixel difference values associated with the TUs may be transformed to produce transform coefficients, which may be quantized.

[0048] A leaf-CU may include one or more prediction units (PUs). In general, a PU represents a spatial area corresponding to all or a portion of the corresponding CU, and may include data for retrieving a reference sample for the PU. Moreover, a PU includes data related to prediction. For example, when the PU is intra-mode encoded, data for the PU may be included in a residual quadtree (RQT), which may include data describing an intra-prediction mode for a TU corresponding to the PU. As another example, when the PU is inter-mode encoded, the PU may include data defining one or more motion vectors for the PU. The data defining the motion vector for a PU may describe, for example, a horizontal component of the motion vector, a vertical component of the motion vector, a resolution for the motion vector (e.g., one-quarter pixel precision or one-eighth pixel precision), a reference picture to which the motion vector points, and/or a reference picture list (e.g., List 0, List 1, or List C) for the motion vector.

[0049] A leaf-CU having one or more PUs may also include one or more transform units (TUs). The transform units may be specified using an RQT (also referred to as a TU quadtree structure), as discussed above. For example, a split flag may indicate whether a leaf-CU is split into four transform units. Then, each transform unit may be split further into further sub-TUs. When a TU is not split further, it may be referred to as a leaf-TU. Generally, for intra coding, all the leaf-TUs belonging to a leaf-CU share the same intra prediction mode. That is, the same intra-prediction mode

is generally applied to calculate predicted values for all TUs of a leaf-CU. For intra coding, a video encoder may calculate a residual value for each leaf-TU using the intra prediction mode, as a difference between the portion of the CU corresponding to the TU and the original block. A TU is not necessarily limited to the size of a PU. Thus, TUs may be larger or smaller than a PU. For intra coding, a PU may be collocated with a corresponding leaf-TU for the same CU. In some examples, the maximum size of a leaf-TU may correspond to the size of the corresponding leaf-CU.

[0050] Moreover, TUs of leaf-CUs may also be associated with respective quadtree data structures, referred to as residual quadtrees (RQTs). That is, a leaf-CU may include a quadtree indicating how the leaf-CU is partitioned into TUs. The root node of a TU quadtree generally corresponds to a leaf-CU, while the root node of a CU quadtree generally corresponds to a treeblock (or LCU). TUs of the RQT that are not split are referred to as leaf-TUs. In general, this disclosure uses the terms CU and TU to refer to leaf-CU and leaf-TU, respectively, unless noted otherwise.

[0051] A video sequence typically includes a series of video frames or pictures. A group of pictures (GOP) generally comprises a series of one or more of the video pictures. A GOP may include syntax data in a header of the GOP, a header of one or more of the pictures, or elsewhere, that describes a number of pictures included in the GOP. Each slice of a picture may include slice syntax data that describes an encoding mode for the respective slice. Video encoder 20 typically operates on video blocks within individual video slices in order to encode the video data. A video block may correspond to a coding node within a CU. The video blocks may have fixed or varying sizes, and may differ in size according to a specified coding standard.

[0052] As an example, the HM supports prediction in various PU sizes. Assuming that the size of a particular CU is 2Nx2N, the HM supports intra-prediction in PU sizes of 2Nx2N or NxN, and inter-prediction in symmetric PU sizes of 2Nx2N, 2NxN, Nx2N, or NxN. The HM also supports asymmetric partitioning for inter-prediction in PU sizes of 2NxnU, 2NxnD, nLx2N, and nRx2N. In asymmetric partitioning, one direction of a CU is not partitioned, while the other direction is partitioned into 25% and 75%. The portion of the CU corresponding to the 25% partition is indicated by an "n" followed by an indication of "Up", "Down," "Left," or "Right." Thus, for example, "2NxnU" refers to a 2Nx2N CU that is partitioned horizontally with a 2Nx0.5N PU on top and a 2Nx1.5N PU on bottom.

[0053] In this disclosure, "NxN" and "N by N" may be used interchangeably to refer to the pixel dimensions of a video block in terms of vertical and horizontal dimensions, e.g., 16x16 pixels or 16 by 16 pixels. In general, a 16x16 block will have 16 pixels in a vertical direction (y = 16) and 16 pixels in a horizontal direction (x = 16). Likewise, an NxN block generally has N pixels in a vertical direction and N pixels in a horizontal direction, where N represents a nonnegative integer value. The pixels in a block may be arranged in rows and columns. Moreover, blocks need not necessarily have the same number of pixels in the horizontal direction as in the vertical direction. For example, blocks may comprise NxM pixels, where M is not necessarily equal to N.

[0054] Following intra-predictive or inter-predictive coding using the PUs of a CU, video encoder 20 may calculate residual data for the TUs of the CU. The PUs may comprise syntax data describing a method or mode of generating predictive pixel data in the spatial domain (also referred to as the pixel domain) and the TUs may comprise coefficients in the transform domain following application of a transform, e.g., a discrete sine transform (DST), a discrete cosine transform (DCT), an integer transform, a wavelet transform, or a conceptually similar transform to residual video data. The residual data may correspond to pixel differences between pixels of the unencoded picture and prediction values corresponding to the PUs. Video encoder 20 may form the TUs including the residual data for the CU, and then transform the TUs to produce transform coefficients for the CU.

[0055] Following any transforms to produce transform coefficients, video encoder 20 may perform quantization of the transform coefficients. Quantization is a broad term intended to have its broadest ordinary meaning. In one embodiment, quantization refers to a process in which transform coefficients are quantized to possibly reduce the amount of data used to represent the coefficients, providing further compression. The quantization process may reduce the bit depth associated with some or all of the coefficients. For example, an n-bit value may be rounded down to an m-bit value during quantization, where n is greater than m.

[0056] Following quantization, the video encoder may scan the transform coefficients, producing a one-dimensional vector from the two-dimensional matrix including the quantized transform coefficients. The scan may be designed to place higher energy (and therefore lower frequency) coefficients at the front of the array and to place lower energy (and therefore higher frequency) coefficients at the back of the

array. In some examples, video encoder 20 may utilize a predefined scan order to scan the quantized transform coefficients to produce a serialized vector that can be entropy encoded. In other examples, video encoder 20 may perform an adaptive scan. After scanning the quantized transform coefficients to form a one-dimensional vector, video encoder 20 may entropy encode the one-dimensional vector, e.g., according to context-adaptive variable length coding (CAVLC), context-adaptive binary arithmetic coding (CABAC), syntax-based context-adaptive binary arithmetic coding (SBAC), Probability Interval Partitioning Entropy (PIPE) coding or another entropy encoding methodology. Video encoder 20 may also entropy encode syntax elements associated with the encoded video data for use by video decoder 30 in decoding the video data.

[0057] To perform CABAC, video encoder 20 may assign a context within a context model to a symbol to be transmitted. The context may relate to, for example, whether neighboring values of the symbol are non-zero or not. To perform CAVLC, video encoder 20 may select a variable length code for a symbol to be transmitted. Codewords in VLC may be constructed such that relatively shorter codes correspond to more probable symbols, while longer codes correspond to less probable symbols. In this way, the use of VLC may achieve a bit savings over, for example, using equal-length codewords for each symbol to be transmitted. The probability determination may be based on a context assigned to the symbol.

[0058] Video encoder 20 may further send syntax data, such as block-based syntax data, frame-based syntax data, and GOP-based syntax data, to video decoder 30, e.g., in a frame header, a block header, a slice header, or a GOP header. The GOP syntax data may describe a number of frames in the respective GOP, and the frame syntax data may indicate an encoding/prediction mode used to encode the corresponding frame.

Video Encoder

[0059] FIG. 2A is a block diagram illustrating an example of a video encoder that may implement techniques in accordance with aspects described in this disclosure. Video encoder 20 may be configured to process a single layer of a video bitstream, such as for HEVC. Further, video encoder 20 may be configured to perform any or all of the techniques of this disclosure, including but not limited to the methods of independent control of inter-layer motion prediction reference resampling and inter-layer sample

prediction reference resampling, methods of processing bitstream constraints relating to inter-layer prediction types, and related processes described in greater detail above and below with respect to FIGS. 4-5. As one example, inter-layer prediction unit 66 (when provided) may be configured to perform any or all of the techniques described in this disclosure. However, aspects of this disclosure are not so limited. In some examples, the techniques described in this disclosure may be shared among the various components of video encoder 20. In some examples, additionally or alternatively, a processor (not shown) may be configured to perform any or all of the techniques described in this disclosure.

[0060] For purposes of explanation, this disclosure describes video encoder 20 in the context of HEVC coding. However, the techniques of this disclosure may be applicable to other coding standards or methods. The encoder 20 of FIG. 2A illustrates a single layer of a codec. However, as will be described further with respect to FIG. 2B, some or all of the video encoder 20 may be duplicated for processing according to a multi-layer codec.

[0061] Video encoder 20 may perform intra-, inter-, and inter-layer prediction (sometime referred to as intra-, inter- or inter-layer coding) of video blocks within video slices. Intra coding relies on spatial prediction to reduce or remove spatial redundancy in video within a given video frame or picture. Inter-coding relies on temporal prediction to reduce or remove temporal redundancy in video within adjacent frames or pictures of a video sequence. Inter-layer coding relies on prediction based upon video within a different layer(s) within the same video coding sequence. Intra-mode (I mode) may refer to any of several spatial based coding modes. Inter-modes, such as uni-directional prediction (P mode) or bi-prediction (B mode), may refer to any of several temporal-based coding modes.

[0062] As shown in FIG. 2A, video encoder 20 receives a current video block within a video frame to be encoded. In the example of FIG. 2A, video encoder 20 includes mode select unit 40, reference frame memory 64, summer 50, transform processing unit 52, quantization unit 54, and entropy encoding unit 56. Mode select unit 40, in turn, includes motion compensation unit 44, motion estimation unit 42, intra-prediction unit 46, inter-layer prediction unit 66, and partition unit 48. Reference frame memory 64 may include a decoded picture buffer. The decoded picture buffer is a broad

term having its ordinary meaning, and in some embodiments refers to a video codec-managed data structure of reference frames.

[0063] For video block reconstruction, video encoder 20 also includes inverse quantization unit 58, inverse transform unit 60, and summer 62. A deblocking filter (not shown in FIG. 2A) may also be included to filter block boundaries to remove blockiness artifacts from reconstructed video. If desired, the deblocking filter would typically filter the output of summer 62. Additional filters (in loop or post loop) may also be used in addition to the deblocking filter. Such filters are not shown for brevity, but if desired, may filter the output of summer 50 (as an in-loop filter).

[0064] During the encoding process, video encoder 20 receives a video frame or slice to be coded. The frame or slice may be divided into multiple video blocks. Motion estimation unit 42 and motion compensation unit 44 perform inter-predictive coding of the received video block relative to one or more blocks in one or more reference frames to provide temporal prediction. Intra-prediction unit 46 may alternatively perform intra-predictive coding of the received video block relative to one or more neighboring blocks in the same frame or slice as the block to be coded to provide spatial prediction. Video encoder 20 may perform multiple coding passes, e.g., to select an appropriate coding mode for each block of video data.

[0065] Moreover, partition unit 48 may partition blocks of video data into sub-blocks, based on evaluation of previous partitioning schemes in previous coding passes. For example, partition unit 48 may initially partition a frame or slice into LCUs, and partition each of the LCUs into sub-CUs based on rate-distortion analysis (e.g., rate-distortion optimization, etc.). Mode select unit 40 may further produce a quadtree data structure indicative of partitioning of an LCU into sub-CUs. Leaf-node CUs of the quadtree may include one or more PUs and one or more TUs.

[0066] Mode select unit 40 may select one of the coding modes, intra, inter, or inter-layer prediction mode, e.g., based on error results, and provide the resulting intra-, inter-, or inter-layer coded block to summer 50 to generate residual block data and to summer 62 to reconstruct the encoded block for use as a reference frame. Mode select unit 40 also provides syntax elements, such as motion vectors, intra-mode indicators, partition information, and other such syntax information, to entropy encoding unit 56.

[0067] Motion estimation unit 42 and motion compensation unit 44 may be highly integrated, but are illustrated separately for conceptual purposes. Motion estimation, performed by motion estimation unit 42, is the process of generating motion vectors, which estimate motion for video blocks. A motion vector, for example, may indicate the displacement of a PU of a video block within a current video frame or picture relative to a predictive block within a reference frame (or other coded unit) relative to the current block being coded within the current frame (or other coded unit). A predictive block is a block that is found to closely match the block to be coded, in terms of pixel difference, which may be determined by sum of absolute difference (SAD), sum of square difference (SSD), or other difference metrics. In some examples, video encoder 20 may calculate values for sub-integer pixel positions of reference pictures stored in reference frame memory 64. For example, video encoder 20 may interpolate values of one-quarter pixel positions, one-eighth pixel positions, or other fractional pixel positions of the reference picture. Therefore, motion estimation unit 42 may perform a motion search relative to the full pixel positions and fractional pixel positions and output a motion vector with fractional pixel precision.

[0068] Motion estimation unit 42 calculates a motion vector for a PU of a video block in an inter-coded slice by comparing the position of the PU to the position of a predictive block of a reference picture. The reference picture may be selected from a first reference picture list (List 0) or a second reference picture list (List 1), each of which identify one or more reference pictures stored in reference frame memory 64. Motion estimation unit 42 sends the calculated motion vector to entropy encoding unit 56 and motion compensation unit 44.

[0069] Motion compensation, performed by motion compensation unit 44, may involve fetching or generating the predictive block based on the motion vector determined by motion estimation unit 42. Motion estimation unit 42 and motion compensation unit 44 may be functionally integrated, in some examples. Upon receiving the motion vector for the PU of the current video block, motion compensation unit 44 may locate the predictive block to which the motion vector points in one of the reference picture lists. Summer 50 forms a residual video block by subtracting pixel values of the predictive block from the pixel values of the current video block being coded, forming pixel difference values, as discussed below. In some embodiments, motion estimation unit 42 can perform motion estimation relative to luma components,

and motion compensation unit 44 can use motion vectors calculated based on the luma components for both chroma components and luma components. Mode select unit 40 may generate syntax elements associated with the video blocks and the video slice for use by video decoder 30 in decoding the video blocks of the video slice.

[0070] Intra-prediction unit 46 may intra-predict or calculate a current block, as an alternative to the inter-prediction performed by motion estimation unit 42 and motion compensation unit 44, as described above. In particular, intra-prediction unit 46 may determine an intra-prediction mode to use to encode a current block. In some examples, intra-prediction unit 46 may encode a current block using various intra-prediction modes, e.g., during separate encoding passes, and intra-prediction unit 46 (or mode select unit 40, in some examples) may select an appropriate intra-prediction mode to use from the tested modes.

[0071] For example, intra-prediction unit 46 may calculate rate-distortion values using a rate-distortion analysis for the various tested intra-prediction modes, and select the intra-prediction mode having the best rate-distortion characteristics among the tested modes. Rate-distortion analysis generally determines an amount of distortion (or error) between an encoded block and an original, unencoded block that was encoded to produce the encoded block, as well as a bitrate (that is, a number of bits) used to produce the encoded block. Intra-prediction unit 46 may calculate ratios from the distortions and rates for the various encoded blocks to determine which intra-prediction mode exhibits the best rate-distortion value for the block.

[0072] After selecting an intra-prediction mode for a block, intra-prediction unit 46 may provide information indicative of the selected intra-prediction mode for the block to entropy encoding unit 56. Entropy encoding unit 56 may encode the information indicating the selected intra-prediction mode. Video encoder 20 may include in the transmitted bitstream configuration data, which may include a plurality of intra-prediction mode index tables and a plurality of modified intra-prediction mode index tables (also referred to as codeword mapping tables), definitions of encoding contexts for various blocks, and indications of a most probable intra-prediction mode, an intra-prediction mode index table, and a modified intra-prediction mode index table to use for each of the contexts.

[0073] The video encoder 20 may include an inter-layer prediction unit 66. Inter-layer prediction unit 66 is configured to predict a current block (e.g., a current

block in the EL) using one or more different layers that are available in SVC (e.g., a base or reference layer). Such prediction may be referred to as inter-layer prediction. Inter-layer prediction unit 66 utilizes prediction methods to reduce inter-layer redundancy, thereby improving coding efficiency and reducing computational resource requirements. Some examples of inter-layer prediction include inter-layer intra prediction, inter-layer motion prediction, and inter-layer residual prediction. Inter-layer intra prediction uses the reconstruction of co-located blocks in the base layer to predict the current block in the enhancement layer. Inter-layer motion prediction uses motion information of the base layer to predict motion in the enhancement layer. Inter-layer residual prediction uses the residue of the base layer to predict the residue of the enhancement layer. When the base and enhancement layers have different spatial resolutions, spatial motion vector scaling and/or inter-layer position mapping using a temporal scaling function may be performed by the inter-layer prediction unit 66, as described in greater detail below.

[0074] Video encoder 20 forms a residual video block by subtracting the prediction data from mode select unit 40 from the original video block being coded. Summer 50 represents the component or components that perform this subtraction operation. Transform processing unit 52 applies a transform, such as a discrete cosine transform (DCT) or a conceptually similar transform, to the residual block, producing a video block comprising residual transform coefficient values. Transform processing unit 52 may perform other transforms which are conceptually similar to DCT. For example, discrete sine transforms (DST), wavelet transforms, integer transforms, sub-band transforms or other types of transforms can also be used.

[0075] Transform processing unit 52 can apply the transform to the residual block, producing a block of residual transform coefficients. The transform may convert the residual information from a pixel value domain to a transform domain, such as a frequency domain. Transform processing unit 52 may send the resulting transform coefficients to quantization unit 54. Quantization unit 54 quantizes the transform coefficients to further reduce bit rate. The quantization process may reduce the bit depth associated with some or all of the coefficients. The degree of quantization may be modified by adjusting a quantization parameter. In some examples, quantization unit 54 may then perform a scan of the matrix including the quantized transform coefficients. Alternatively, entropy encoding unit 56 may perform the scan.

[0076] Following quantization, entropy encoding unit 56 entropy encodes the quantized transform coefficients. For example, entropy encoding unit 56 may perform context adaptive variable length coding (CAVLC), context adaptive binary arithmetic coding (CABAC), syntax-based context-adaptive binary arithmetic coding (SBAC), probability interval partitioning entropy (PIPE) coding or another entropy coding technique. In the case of context-based entropy coding, context may be based on neighboring blocks. Following the entropy coding by entropy encoding unit 56, the encoded bitstream may be transmitted to another device (e.g., video decoder 30) or archived for later transmission or retrieval.

[0077] Inverse quantization unit 58 and inverse transform unit 60 apply inverse quantization and inverse transformation, respectively, to reconstruct the residual block in the pixel domain (e.g., for later use as a reference block). Motion compensation unit 44 may calculate a reference block by adding the residual block to a predictive block of one of the frames of reference frame memory 64. Motion compensation unit 44 may also apply one or more interpolation filters to the reconstructed residual block to calculate sub-integer pixel values for use in motion estimation. Summer 62 adds the reconstructed residual block to the motion compensated prediction block produced by motion compensation unit 44 to produce a reconstructed video block for storage in reference frame memory 64. The reconstructed video block may be used by motion estimation unit 42 and motion compensation unit 44 as a reference block to inter-code a block in a subsequent video frame.

Multi-Layer Video Encoder

[0078] FIG. 2B is a block diagram illustrating an example of a multi-layer video encoder 21 that may implement techniques in accordance with aspects described in this disclosure. The video encoder 21 may be configured to process multi-layer video frames, such as for SHVC and multiview coding. Further, the video encoder 21 may be configured to perform any or all of the techniques of this disclosure.

[0079] The video encoder 21 includes a video encoder 20A and video encoder 20B, each of which may be configured as the video encoder 20 of FIG. 2A and may perform the functions described above with respect to the video encoder 20. Further, as indicated by the reuse of reference numbers, the video encoders 20A and 20B may include at least some of the systems and subsystems as the video encoder 20.

Although the video encoder 21 is illustrated as including two video encoders 20A and 20B, the video encoder 21 is not limited as such and may include any number of video encoder 20 layers. In some embodiments, the video encoder 21 may include a video encoder 20 for each picture or frame in an access unit. For example, an access unit that includes five pictures may be processed or encoded by a video encoder that includes five encoder layers. In some embodiments, the video encoder 21 may include more encoder layers than frames in an access unit. In some such cases, some of the video encoder layers may be inactive when processing some access units.

[0080] In addition to the video encoders 20A and 20B, the video encoder 21 may include a resampling unit 90. The resampling unit 90 may, in some cases, upsample a base layer of a received video frame to, for example, create an enhancement layer. The resampling unit 90 may upsample particular information associated with the received base layer of a frame, but not other information. For example, the resampling unit 90 may upsample the spatial size or number of pixels of the base layer, but the number of slices or the picture order count may remain constant. In some cases, the resampling unit 90 may not process the received video and/or may be optional. For example, in some cases, the mode select unit 40 may perform upsampling. In some embodiments, the resampling unit 90 is configured to upsample a layer and reorganize, redefine, modify, or adjust one or more slices to comply with a set of slice boundary rules and/or raster scan rules. Although primarily described as upsampling a base layer, or a lower layer in an access unit, in some cases, the resampling unit 90 may downsample a layer. For example, if during streaming of a video bandwidth is reduced, a frame may be downsampled instead of upsampled. Resampling unit 90 may be further configured to perform cropping and/or padding operations, as well.

[0081] The resampling unit 90 may be configured to receive a picture or frame (or picture information associated with the picture) from the decoded picture buffer 114 of the lower layer encoder (e.g., the video encoder 20A) and to upsample the picture (or the received picture information). This upsampled picture may then be provided to the mode select unit 40 of a higher layer encoder (e.g., the video encoder 20B) configured to encode a picture in the same access unit as the lower layer encoder. In some cases, the higher layer encoder is one layer removed from the lower layer encoder. In other cases, there may be one or more higher layer encoders between the layer 0 video encoder and the layer 1 encoder of FIG. 2B.

[0082] In some cases, the resampling unit 90 may be omitted or bypassed. In such cases, the picture from the decoded picture buffer 64 of the video encoder 20A may be provided directly, or at least without being provided to the resampling unit 90, to the mode select unit 40 of the video encoder 20B. For example, if video data provided to the video encoder 20B and the reference picture from the decoded picture buffer 64 of the video encoder 20A are of the same size or resolution, the reference picture may be provided to the video encoder 20B without any resampling.

[0083] In some embodiments, the video encoder 21 downsamples video data to be provided to the lower layer encoder using the downsampling unit 94 before provided the video data to the video encoder 20A. Alternatively, the downsampling unit 94 may be a resampling unit 90 capable of upsampling or downsampling the video data. In yet other embodiments, the downsampling unit 94 may be omitted.

[0084] As illustrated in FIG. 2B, the video encoder 21 may further include a multiplexor 98, or mux. The mux 98 can output a combined bitstream from the video encoder 21. The combined bitstream may be created by taking a bitstream from each of the video encoders 20A and 20B and alternating which bitstream is output at a given time. While in some cases the bits from the two (or more in the case of more than two video encoder layers) bitstreams may be alternated one bit at a time, in many cases the bitstreams are combined differently. For example, the output bitstream may be created by alternating the selected bitstream one block at a time. In another example, the output bitstream may be created by outputting a non-1:1 ratio of blocks from each of the video encoders 20A and 20B. For instance, two blocks may be output from the video encoder 20B for each block output from the video encoder 20A. In some embodiments, the output stream from the mux 98 may be preprogrammed. In other embodiments, the mux 98 may combine the bitstreams from the video encoders 20A, 20B based on a control signal received from a system external to the video encoder 21, such as from a processor on the source device 12. The control signal may be generated based on the resolution or bitrate of a video from the video source 18, based on a bandwidth of the channel 16, based on a subscription associated with a user (e.g., a paid subscription versus a free subscription), or based on any other factor for determining a resolution output desired from the video encoder 21.

Video Decoder

[0085] FIG. 3A is a block diagram illustrating an example of a video decoder that may implement techniques in accordance with aspects described in this disclosure. The video decoder 30 may be configured to process a single layer of a video bitstream, such as for HEVC. Further, video decoder 30 may be configured to perform any or all of the techniques of this disclosure, including but not limited to the methods of independent control of inter-layer motion prediction reference resampling and inter-layer sample prediction reference resampling, methods of processing bitstream constraints relating to inter-layer prediction types, and related processes described in greater detail above and below with respect to FIGS. 4-5. As one example, inter-layer prediction unit 75 may be configured to perform any or all of the techniques described in this disclosure. However, aspects of this disclosure are not so limited. In some examples, the techniques described in this disclosure may be shared among the various components of video decoder 30. In some examples, additionally or alternatively, a processor (not shown) may be configured to perform any or all of the techniques described in this disclosure.

[0086] For purposes of explanation, this disclosure describes video decoder 30 in the context of HEVC coding. However, the techniques of this disclosure may be applicable to other coding standards or methods. The decoder 30 of FIG. 3A illustrates a single layer of a codec. However, as will be described further with respect to FIG. 3B, some or all of the video decoder 30 may be duplicated for processing according to a multi-layer codec.

[0087] In the example of FIG. 3A, video decoder 30 includes an entropy decoding unit 70, motion compensation unit 72, intra prediction unit 74, inter-layer prediction unit 75, inverse quantization unit 76, inverse transformation unit 78, reference frame memory 82 and summer 80. In some embodiments, motion compensation unit 72 and/or intra prediction unit 74 may be configured to perform inter-layer prediction, in which case the inter-layer prediction unit 75 may be omitted. Video decoder 30 may, in some examples, perform a decoding pass generally reciprocal to the encoding pass described with respect to video encoder 20 (FIG. 2A). Motion compensation unit 72 may generate prediction data based on motion vectors received from entropy decoding unit 70, while intra-prediction unit 74 may generate prediction data based on intra-prediction mode indicators received from entropy decoding unit 70.

Reference frame memory 82 may include a decoded picture buffer. The decoded picture buffer is a broad term having its ordinary meaning, and in some embodiments refers to a video codec-managed data structure of reference frames.

[0088] During the decoding process, video decoder 30 receives an encoded video bitstream that represents video blocks of an encoded video slice and associated syntax elements from video encoder 20. Entropy decoding unit 70 of video decoder 30 entropy decodes the bitstream to generate quantized coefficients, motion vectors or intra-prediction mode indicators, and other syntax elements. Entropy decoding unit 70 forwards the motion vectors to and other syntax elements to motion compensation unit 72. Video decoder 30 may receive the syntax elements at the video slice level and/or the video block level.

[0089] When the video slice is coded as an intra-coded (I) slice, intra prediction unit 74 may generate prediction data for a video block of the current video slice based on a signaled intra prediction mode and data from previously decoded blocks of the current frame or picture. When the video frame is coded as an inter-coded (e.g., B, P or GPB) slice, motion compensation unit 72 produces predictive blocks for a video block of the current video slice based on the motion vectors and other syntax elements received from entropy decoding unit 70. The predictive blocks may be produced from one of the reference pictures within one of the reference picture lists. Video decoder 30 may construct the reference frame lists, List 0 and List 1, using default construction techniques based on reference pictures stored in reference frame memory 82. Motion compensation unit 72 determines prediction information for a video block of the current video slice by parsing the motion vectors and other syntax elements, and uses the prediction information to produce the predictive blocks for the current video block being decoded. For example, motion compensation unit 72 uses some of the received syntax elements to determine a prediction mode (e.g., intra- or inter-prediction) used to code the video blocks of the video slice, an inter-prediction slice type (e.g., B slice, P slice, or GPB slice), construction information for one or more of the reference picture lists for the slice, motion vectors for each inter-encoded video block of the slice, inter-prediction status for each inter-coded video block of the slice, and other information to decode the video blocks in the current video slice.

[0090] Motion compensation unit 72 may also perform interpolation based on interpolation filters. Motion compensation unit 72 may use interpolation filters as

used by video encoder 20 during encoding of the video blocks to calculate interpolated values for sub-integer pixels of reference blocks. In this case, motion compensation unit 72 may determine the interpolation filters used by video encoder 20 from the received syntax elements and use the interpolation filters to produce predictive blocks.

[0091] Video decoder 30 may also include an inter-layer prediction unit 75. The inter-layer prediction unit 75 is configured to predict a current block (e.g., a current block in the EL) using one or more different layers that are available in SVC (e.g., a base or reference layer). Such prediction may be referred to as inter-layer prediction. Inter-layer prediction unit 75 utilizes prediction methods to reduce inter-layer redundancy, thereby improving coding efficiency and reducing computational resource requirements. Some examples of inter-layer prediction include inter-layer intra prediction, inter-layer motion prediction, and inter-layer residual prediction. Inter-layer intra prediction uses the reconstruction of co-located blocks in the base layer to predict the current block in the enhancement layer. Inter-layer motion prediction uses motion information of the base layer to predict motion in the enhancement layer. Inter-layer residual prediction uses the residue of the base layer to predict the residue of the enhancement layer. When the base and enhancement layers have different spatial resolutions, spatial motion vector scaling and/or inter-layer position mapping may be performed by the inter-layer prediction unit 75 using a temporal scaling function, as described in greater detail below.

[0092] Inverse quantization unit 76 inverse quantizes, e.g., de-quantizes, the quantized transform coefficients provided in the bitstream and decoded by entropy decoding unit 70. The inverse quantization process may include use of a quantization parameter QPY calculated by video decoder 30 for each video block in the video slice to determine a degree of quantization and, likewise, a degree of inverse quantization that should be applied.

[0093] Inverse transform unit 78 applies an inverse transform, e.g., an inverse DCT, an inverse DST, an inverse integer transform, or a conceptually similar inverse transform process, to the transform coefficients in order to produce residual blocks in the pixel domain.

[0094] After motion compensation unit 72 generates the predictive block for the current video block based on the motion vectors and other syntax elements, video decoder 30 forms a decoded video block by summing the residual blocks from inverse

transform unit 78 with the corresponding predictive blocks generated by motion compensation unit 72. Summer 90 represents the component or components that perform this summation operation. If desired, a deblocking filter may also be applied to filter the decoded blocks in order to remove blockiness artifacts. Other loop filters (either in the coding loop or after the coding loop) may also be used to smooth pixel transitions, or otherwise improve the video quality. The decoded video blocks in a given frame or picture are then stored in reference frame memory 82, which stores reference pictures used for subsequent motion compensation. Reference frame memory 82 also stores decoded video for later presentation on a display device, such as display device 32 of FIG. 1.

Multi-Layer Decoder

[0095] FIG. 3B is a block diagram illustrating an example of a multi-layer video decoder 31 that may implement techniques in accordance with aspects described in this disclosure. The video decoder 31 may be configured to process multi-layer video frames, such as for SHVC and multiview coding. Further, the video decoder 31 may be configured to perform any or all of the techniques of this disclosure.

[0096] The video decoder 31 includes a video decoder 30A and video decoder 30B, each of which may be configured as the video decoder 30 of FIG. 3A and may perform the functions described above with respect to the video decoder 30. Further, as indicated by the reuse of reference numbers, the video decoders 30A and 30B may include at least some of the systems and subsystems as the video decoder 30. Although the video decoder 31 is illustrated as including two video decoders 30A and 30B, the video decoder 31 is not limited as such and may include any number of video decoder 30 layers. In some embodiments, the video decoder 31 may include a video decoder 30 for each picture or frame in an access unit. For example, an access unit that includes five pictures may be processed or decoded by a video decoder that includes five decoder layers. In some embodiments, the video decoder 31 may include more decoder layers than frames in an access unit. In some such cases, some of the video decoder layers may be inactive when processing some access units.

[0097] In addition to the video decoders 30A and 30B, the video decoder 31 may include an upsampling unit 92. In some embodiments, the upsampling unit 92 may upsample a base layer of a received video frame to create an enhanced layer to be added

to the reference picture list for the frame or access unit. This enhanced layer can be stored in the reference frame memory 82 (e.g., in its decoded picture buffer, etc.). In some embodiments, the upsampling unit 92 can include some or all of the embodiments described with respect to the resampling unit 90 of FIG. 2A. In some embodiments, the upsampling unit 92 is configured to upsample a layer and reorganize, redefine, modify, or adjust one or more slices to comply with a set of slice boundary rules and/or raster scan rules. In some cases, the upsampling unit 92 may be a resampling unit configured to upsample and/or downsample a layer of a received video frame

[0098] The upsampling unit 92 may be configured to receive a picture or frame (or picture information associated with the picture) from the decoded picture buffer 82 of the lower layer decoder (e.g., the video decoder 30A) and to upsample the picture (or the received picture information). This upsampled picture may then be provided to the mode select unit 71 of a higher layer decoder (e.g., the video decoder 30B) configured to decode a picture in the same access unit as the lower layer decoder. In some cases, the higher layer decoder is one layer removed from the lower layer decoder. In other cases, there may be one or more higher layer decoders between the layer 0 decoder and the layer 1 decoder of FIG. 3B.

[0099] In some cases, the upsampling unit 92 may be omitted or bypassed. In such cases, the picture from the decoded picture buffer 82 of the video decoder 30A may be provided directly, or at least without being provided to the upsampling unit 92, to the mode select unit 71 of the video decoder 30B. For example, if video data provided to the video decoder 30B and the reference picture from the decoded picture buffer 82 of the video decoder 30A are of the same size or resolution, the reference picture may be provided to the video decoder 30B without upsampling. Further, in some embodiments, the upsampling unit 92 may be a resampling unit 90 configured to upsample or downsample a reference picture received from the decoded picture buffer 82 of the video decoder 30A.

[00100] As illustrated in FIG. 3B, the video decoder 31 may further include a demultiplexor 99, or demux. The demux 99 can split an encoded video bitstream into multiple bitstreams with each bitstream output by the demux 99 being provided to a different video decoder 30A and 30B. The multiple bitstreams may be created by receiving a bitstream and each of the video decoders 30A and 30B receives a portion of the bitstream at a given time. While in some cases the bits from the bitstream received

at the demux 99 may be alternated one bit at a time between each of the video decoders (e.g., video decoders 30A and 30B in the example of FIG. 3B), in many cases the bitstream is divided differently. For example, the bitstream may be divided by alternating which video decoder receives the bitstream one block at a time. In another example, the bitstream may be divided by a non-1:1 ratio of blocks to each of the video decoders 30A and 30B. For instance, two blocks may be provided to the video decoder 30B for each block provided to the video decoder 30A. In some embodiments, the division of the bitstream by the demux 99 may be preprogrammed. In other embodiments, the demux 99 may divide the bitstream based on a control signal received from a system external to the video decoder 31, such as from a processor on the destination device 14. The control signal may be generated based on the resolution or bitrate of a video from the input interface 28, based on a bandwidth of the channel 16, based on a subscription associated with a user (e.g., a paid subscription versus a free subscription), or based on any other factor for determining a resolution obtainable by the video decoder 31.

Reference Layer Types

[00101] In an implementation of MV-HEVC and SHVC, there is a direct_dependency_flag syntax element specifying what layer can be used for inter-layer prediction. **direct_dependency_flag**[i][j] equal to 0 specifies that the layer with index j is not a direct reference layer for the layer with index i. direct_dependency_flag[i][j] equal to 1 specifies that the layer with index j may be a direct reference layer for the layer with index i. When direct_dependency_flag[i][j] is not present for i and j in the range of 0 to vps_max_layers_minus1, it is inferred to be equal to 0.

[00102] In addition, two types of inter-layer prediction can be applied: inter-layer motion prediction, inter-layer sample prediction, or both. To specify what inter-layer prediction types are available for some particular layer, the direct_dependency_type is signaled.

[00103] **direct_dependency_type**[i][j] is used to derive the variables NumSamplePredRefLayers[i], NumMotionPredRefLayers[i], SamplePredEnabledFlag[i][j], and MotionPredEnabledFlag[i][j]. The variable NumSamplePredRefLayers[i] can refer to the number of reference layers that can be

used for sample prediction for the layer with index i. The variable NumMotionPredRefLayers[i] can refer to the number of reference layers that can be used for motion prediction for the layer with index i. The variable SamplePredEnabledFlag[i][j] can refer to whether sample prediction using the layer with index j is enabled for the layer with index i. The variable MotionPredEnabledFlag[i][j] can refer to motion prediction using the layer with index j is enabled for the layer with index i. direct_dependency_type[i][j] should be in the range of 0 to 2, inclusive, in bitstreams. Although the value of direct_dependency_type[i][j] should be in the range of 0 to 2, inclusive, decoders should allow values of direct_dependency_type[i][j] in the range of 3 to 2^{32}-2, inclusive, to appear in the syntax.

[00104] The variables NumSamplePredRefLayers[i], NumMotionPredRefLayers[i], SamplePredEnabledFlag[i][j], MotionPredEnabledFlag[i][j], NumDirectRefLayers[i], RefLayerId[i][j], MotionPredRefLayerId[i][j], and SamplePredRefLayerId[i][j] are derived as follows:

```
for( i = 0; i < 64; i++ ) {
        NumSamplePredRefLayers[ i ] = 0
        NumMotionPredRefLayers[ i ] = 0
        NumDirectRefLayers[ i ] = 0
        for( j = 0; j < 64; j++ ) {
                SamplePredEnabledFlag[ i ][ j ] = 0
                MotionPredEnabledFlag[ i ][ j ] = 0
                RefLayerId[ i ][ j ] = 0
                SamplePredRefLayerId[ i ][ j ] = 0
                MotionPredRefLayerId[ i ][ j ] = 0
        }
}
for( i = 1; i <= vps_max_layers_minus1; i++ ) {
iNuhLId = layer_id_in_nuh[ i ]
for( j = 0; j < i; j++ )
        if( direct_dependency_flag[ i ][ j ] ) {
                RefLayerId[ iNuhLId ][ NumDirectRefLayers[ iNuhLId ]++ ] =
        layer_id_in_nuh[ j ]
                SamplePredEnabledFlag[ iNuhLId ][ j ] = ( (
        direct_dependency_type[ i ][ j ] + 1 ) & 1 )
                NumSamplePredRefLayers[ iNuhLId ] +=
```

```
                SamplePredEnabledFlag[ iNuhLId ][ j ]
                    MotionPredEnabledFlag[ iNuhLId ][ j ] = ( ( (
direct_dependency_type[ i ][ j ] + 1 ) & 2 ) >> 1 )
                    NumMotionPredRefLayers[ iNuhLId ] +=
MotionPredEnabledFlag[ iNuhLId ][ j ]
        }
        }
        for( i = 1, mIdx = 0, sIdx = 0; i <= vps_max_layers_minus1; i++ ) {
            iNuhLId = layer_id_in_nuh[ i ]
            for( j = 0, j < i; j++ ) {
                if( MotionPredEnabledFlag[ iNuhLId ][ j ] )
                    MotionPredRefLayerId[ iNuhLId ][ mIdx++ ] = layer_id_in_nuh[ j ]
                if( SamplePredEnabledFlag[ INuhLid ][ j ] )
                    SamplePredRefLayerId[ iNuhLid ][ sIdx++ ] = layer_id_in_nuh[ j ]
            }
        }
```

Restriction on Number of Pictures Used for Inter-Layer Reference with Resampling

[00105] In one SHVC implementation, the number of inter-layer reference picture that needs to be resampled for decoding of any particular picture is limited to be up to one. Resampling process, for example, is invoked when the reference and enhancement layer have different picture sizes.

[00106] However, having a limit on the number of inter-layer reference picture that is resampled may create problems, such as the following:

- When decoding a current picture, a picture that is only used for inter-layer motion prediction (not for sample prediction) is also resampled when it has a different spatial resolution than the current picture. However, resampling of such a picture may waste computing resource unnecessarily.
- If a picture that is only used for inter-layer motion prediction exists when decoding a current picture, then no other pictures for inter-layer sample prediction can be sample-resampled according to the restriction that the number of pictures used for inter-layer reference with resampling cannot be greater than one. In other words, when such a picture exists, if there is no other lower-layer picture that has the same resolution as the current picture, inter-layer prediction

of samples cannot be used for the current picture even if there is another low-layer picture with a different spatial resolution.

- Bitstream conformance restrictions are missing for pictures of certain direct reference layers that are indicated as not used for inter-layer sample prediction or indicated as not used for inter-layer motion prediction.
- Pictures for inter-layer reference are included into the initial reference picture lists (before reference picture list modification commands) without making a difference between different types of inter-layer predictions indicated for the pictures, which is sub-optimal.
- Coding of collocated_ref_idx signaled in the slice header and the reference index signaled at block (e.g., CU, PU, etc.) level may use unnecessarily more bits.

Inter-Layer Prediction Types in Multi-Layer Video Coding

[00107] In order to address these and other challenges, the techniques according to certain aspects can independently control the number of reference pictures to be resampled for inter-layer motion prediction and the number of reference pictures to be resampled for inter-layer sample prediction. The techniques can also provide and/or process bitstream constraints relating to inter-layer prediction types. More specifically, the techniques can provide and/or process a bitstream constraint that the collocated reference index (e.g., collocated_ref_idx) can only refer to an ILRP used for at least ILMP. The techniques can also provide and/or process a bitstream constraint that reference index (e.g., ref_idx) can only refer to an ILRP used for at least ILSP.

[00108] In this manner, when decoding a current picture, an ILRP only used in ILMP does not need to be sample-resampled. Also, an ILRP only used in ILMP does not have to prevent another ILRP from being used in ILP. For example, an ILRP used in ILSP can be sample-resampled and used in ILP. This can lead to more accurate prediction and more efficient coding. For example, one more type of ILP (e.g., ILSP in the example above) can be used. In addition, computational complexity can be reduced by avoiding invoking unnecessary resampling processes. Certain details relating to the techniques are described below.

Pictures Indicated as Not Used for Inter-Layer Sample Prediction

[00109] According to certain aspects, the techniques can exclude an inter-layer reference picture from the resampling process if it is indicated to be used only for inter-layer motion prediction. For this type of pictures, sample (pixel) information does not need to be stored in the memory.

[00110] In addition, an inter-layer reference picture indicated to be used only for inter-layer motion prediction is not counted as a picture that requires sample-resampling process, since samples will not be used for inter prediction purpose. Consequently, another low-layer picture with a different spatial resolution can be used for inter-layer prediction of samples for the current picture.

[00111] Moreover, a picture that is indicated as not used for inter-layer sample prediction may not be referred to by a reference index signaled at block (e.g., CU, PU, etc.) level. For example, such picture cannot be used for inter-prediction.

[00112] Furthermore, the total number of reference pictures used in reference index signaling can be adjusted, including only the pictures that can be used for inter prediction, such that the reference index signaled in block (e.g., CU, PU, etc.) level can use fewer bits.

Pictures Indicated as Not Used for Inter-Layer Motion Prediction

[00113] According to certain aspects, for a picture indicated not to be used for inter-layer motion prediction (e.g., indicated to be used only for inter-layer sample prediction), the motion information is not required to be derived and this picture cannot be used for temporal motion vector (TMVP) derivation. For example, such picture cannot be used as a collocated picture in TMVP derivation. And motion information may not be stored for this picture.

[00114] It can be implied that a collocated picture, for example, defined by collocated_ref_idx syntax element, cannot be a picture that is indicated not used for inter-layer motion prediction. In other words, collocated_red_idx should not point to a lower-layer picture that is only for inter-layer sample prediction or that is not used for inter-layer prediction at all.

[00115] In addition, the total number of reference pictures used to define the collocated_ref_idx range can be adjusted to include only the pictures that can be used for TMVP derivation, such that the signaling of collocated_ref_idx can use fewer bits.

[00116] Alternative to not using this type of reference picture as a collocated picture, default motion information can be assigned to the inter-layer sample only prediction pictures. The default motion information can include at least prediction mode, motion vectors, reference indices, and reference picture picture order counts (POCs). For example, intra prediction mode, which specifies no motion information, can be assigned for the inter-layer sample only prediction picture. In such case, no TMVP may be derived if this picture is used as a collocated picture due to intra prediction mode being assigned. Accordingly, default motion information may be assigned.

Separate Constraint on Number of Sample Resampled and Motion Resampled Inter-Layer Pictures

[00117] In one embodiment, inter-layer reference pictures used for inter-layer motion prediction are counted separately from the inter-layer sample prediction towards the constraint on the number of the resampled pictures. In early versions of SHVC, only one resampled picture could be used; sample-resampled and motion-resampled pictures were not counted separately, and as a result, only one ILP type (e.g., ILMP only or ILSP only) could be used in some cases, as mentioned above. If the sample-resampled and motion-resampled pictures are counted separately, then up to one sample resampling can be applied, and up to one motion resampling can be applied.

[00118] In another embodiment, the number of sample resampled inter-layer pictures and the number of motion resampled inter-layer pictures can be restricted and/or limited separately with a different number. For example, one sample resampled inter-layer picture and two motion resampled inter-layer pictures can be restricted to be used.

[00119] The above mentioned techniques can be implemented as shown in the following examples. The examples are provided in the context of early versions of SHVC. Changes from the early versions of SHVC are indicated in italics. In certain embodiments, the techniques may not count the picture used only for inter-layer motion prediction towards resampled pictures number.

Example 1

G.8.1.2 Decoding process for inter-layer reference picture set

Output of this process is an updated list of inter-layer reference pictures RefPicSetInterLayer.

The list RefPicSetInterLayer is first emptied and then derived as follows.

for(i = 0; i < NumActiveRefLayerPics; i++) {

if(there is a picture picX in the DPB that is in the same access unit as the current picture and has nuh_layer_id equal to RefPicLayerId[i]) {

If SamplePredEnabledFlag[nuh_layer_id][RefPicLayerId[i]] is equal to 0 the following applies

- the picture motion field resampling process as specified in subclause G.8.1.4.2 is invoked with rlPicMotion as input, and with the resampled motion field of rsPicMotion as output, where the variable rlPicMotion is defined as a group of variable arrays specifying the compressed motion field of rlPic and the variable rsPicMotion is defined as a group of variable arrays specifying the resampled motion field of rsPic.

Otherwise,

- an interlayer reference picture rsPic is derived by invoking the subclause G.8.1.4 with picX given as input

RefPicSetInterLayer[i] = rsPic

RefPicSetInterLayer[i] is marked as "used for long-term reference"

} else

RefPicSetInterLayer[i] = "no reference picture"

}

There should be no entry equal to "no reference picture" in RefPicSetInterLayer.

Table 1 – Example 1

[00120] In Example 1, the motion-resampling process is called separately from the sample-resampling process where ILRP is used for ILMP only. In this example, when ILRP picture is used for both ILMP and ILSP, motion-resampling process is invoked through the sample-resampling process. An alternative description is provided in Example 2. In the examples and description above and below, portions in italics may indicate changes to the early versions of SHVC. Portions in underline may indicate portions that are specific for SHVC only and not present in the MV-HEVC.

Example 2

[00121] In Example 2, an alternative description is provided where the two processes (e.g., motion-resampling and sample-resampling) are called independently depending on whether ILRP is used for ILMP and/or ILSP. The invocation of the motion-resampling process is removed from the sample-resampling in the section G.8.1.4 and moved to a separate section G.8.1.5, for example, in order to improve readability of the specification text.

G.8.1.2 Decoding process for inter-layer reference picture set

Output of this process is an updated list of inter-layer reference pictures RefPicSetInterLayer.

The list RefPicSetInterLayer is first emptied and then derived as follows.

for(i = 0; i < NumActiveRefLayerPics; i++) {

if(there is a picture picX in the DPB that is in the same access unit as the current picture and has

nuh_layer_id equal to RefPicLayerId[i]) {

if(MotionPredEnabledFlag[nuh_layer_id][RefPicLayerId[i]])

a motion resampled field rsPicMotion of an interlayer reference picture rsPic is derived by invoking subclause G.8.1.5 with the compressed motion field of picX as input

if(SamplePredEnabledFlag[nuh_layer_id][RefPicLayerId[i]])

a sample resampled rsPicSample of an interlayer reference picture rsPic is derived by invoking the subclause G.8.1.4 with the samples of picX and DirectRefLayerIdx[currLayerId][RefPicLayerId[i]] as inputs

RefPicSetInterLayer[i] = rsPic

RefPicSetInterLayer[i] is marked as "used for long-term reference"

} else

RefPicSetInterLayer[i] = "no reference picture"

}

There shall be no entry equal to "no reference picture" in RefPicSetInterLayer.

If the current picture is a RADL picture, there shall be no entry in the RefPicSetInterLayer that is a RASL picture.

G.8.1.4 Resampling process for inter layer reference pictures

Input to this process is a decoded reference layer picture rlPic.

Output of this process is the resampled reference layer picture rsPic.

The variables PicWidthInSamplesL and PicHeightInSamplesL are set equal to pic_width_in_luma_samples and pic_height_in_luma_samples, respectively. The variable rsPicSample is defined as a group of sample arrays specifying the resampled sample values of rsPic of the luma and chroma components. The variable rsPicMotion is defined as a group of variable arrays specifying the resampled motion field of rsPic.

The variables RefLayerPicWidthInSamplesL and RefLayerPicHeightInSamplesL are set equal to the width and height of the decoded reference layer picture rlPic in units of luma samples, respectively. The variable rlPicSample is defined as a group of sample arrays specifying the sample values of rlPic of the luma and chroma components. The variable rlPicMotion is defined as a group of variable arrays specifying the compressed motion field of rlPic.

The variables PicWidthInSamplesC, PicHeightInSamplesC, RefLayerPicWidthInSamplesC, and RefLayerPicHeightInSamplesC are derived as follows:

PicWidthInSamplesC = PicWidthInSamplesL / subWidthC

PicHeightInSamplesC = PicHeightInSamplesL / subHeightC (G-11)

RefLayerPicWidthInSamplesC = RefLayerPicWidthInSamplesL / subWidthC (G-12)

RefLayerPicHeightInSamplesC = RefLayerPicHeightInSamplesL / subHeightC (G-13)

The variables ScaledRefLayerLeftOffset, ScaledRefLayerTopOffset, ScaledRefLayerRightOffset and ScaledRefLayerBottomOffset are derived as follows:

ScaledRefLayerLeftOffset = scaled_ref_layer_left_offset << 1 (G-14)

ScaledRefLayerTopOffset = scaled_ref_layer_top_offset << 1 (G-15)

ScaledRefLayerRightOffset = scaled_ref_layer_right_offset << 1 (G-16)

ScaledRefLayerBottomOffset = scaled_ref_layer_bottom_offset << 1 (G-17)

The variables ScaledRefLayerPicWidthInSamplesL and ScaledRefLayerPicHeightInSamplesL are derived as follows:

ScaledRefLayerPicWidthInSamplesL = PicWidthInSamplesL –
ScaledRefLayerLeftOffset – ScaledRefLayerRightOffset (G-18)

ScaledRefLayerPicHeightInSamplesL = PicHeightInSamplesL –
ScaledRefLayerTopOffset – ScaledRefLayerBottomOffset (G-19)

The variables ScaleFactorX and ScaleFactorY are derived as follows:

ScaleFactorX = ((RefLayerPicWidthInSamplesL
<< 16) + (ScaledRefLayerPicWidthInSamplesL >> 1)) /
Scaled RefLayerPic WidthInSamplesL (G-20)

ScaleFactorY = ((RefLayerPicHeightInSamplesL
<< 16) + (ScaledRefLayerPicHeightInSamplesL >> 1)) /
ScaledRefLayer PicHeightInSamplesL (G-21)

The following steps are applied to derive the resampled inter layer reference picture rsPic.

– if PicWidthInSamplesL is equal to RefLayerPicWidthInSamplesL and PicHeightInSamplesL is equal to RefLayerPicHeightInSamplesL and the values of ScaledRefLayerLeftOffset, ScaledRefLayerTopOffset, ScaledRefLayerRightOffset and ScaledRefLayerBottomOffset are all equal to 0
  – rsPicSample is set equal to rlPicSample,
  – When alt_collocated_indication_flag is equal to 1, rsPicMotion is set equal to rlPicMotion.
– otherwise, rsPic is derived as follows:
  – The picture sample resampling process as specified in subclause G.8.1.4.1 is invoked with the

sample values of rlPicSample as input, and with the resampled sample values of rsPicSample as output.

~~*When alt_collocated_indication_flag is equal to 1, the picture motion field resampling process as specified in subclause G.8.1.4.2 is invoked with rlPicMotion as input, and with the resampled motion field of rsPicMotion as output.*~~

~~*G.8.1.4.1.1*~~ *G.8.1.5* Resampling process of picture motion field

Input to this process is rlPicMotion specifying the motion field of the picture rlPic.

Output of this process is rsPicMotion specifying the resampled motion field of the resampled picture.

The motion field of rlPic specified by rlPicMotion consists of:

- a (RefLayerPicWidthInSamplesL) x (RefLayerPicHeightInSamplesL) array predModeRL specifies the prediction modes of the reference layer picture rlPic,
- two (RefLayerPicWidthInSamplesL) x (RefLayerPicHeightInSamplesL) arrays refIdxLXRL specify the reference indices of the reference layer picture rlPic, with X = 0,1,
- two (RefLayerPicWidthInSamplesL) x (RefLayerPicHeightInSamplesL) arrays mvLXRL specify the luma motion vectors of the reference layer picture rlPic, with X = 0,1,
- two (RefLayerPicWidthInSamplesL) x (RefLayerPicHeightInSamplesL) arrays refPicOrderCntLXRL specify the reference picture order counts of the reference layer picture rlPic, with X = 0, 1,
- two (RefLayerPicWidthInSamplesL) x (RefLayerPicHeightInSamplesL) arrays predFlagLXRL specify the prediction list utilization flags of the reference layer picture rlPic, with X = 0,1.

The resampled motion field specified by rsPicMotion consists of:

- a (PicWidthInSamplesL) x (PicHeightInSamplesL) array predMode specifies the prediction modes of the resampled picture,
- two (PicWidthInSamplesL) x (PicHeightInSamplesL) arrays refIdxLX specify the reference indexes of the resampled picture, with X = 0,1,
- two (PicWidthInSamplesL) x (PicHeightInSamplesL) arrays mvLX specify the luma motion vectors of the resampled picture, with X = 0,1,
- two (PicWidthInSamplesL) x (PicHeightInSamplesL) arrays refPicOrderCntLX specify the reference picture order counts of the resampled picture, with X = 0, 1
- two (PicWidthInSamplesL) x (PicHeightInSamplesL) arrays predFlagLX specify the prediction list utilization flags of the resampled picture, with X = 0,1.

For each luma sample location xPb = 0 ... ((PicWidthInSamplesL + 15) >> 4) − 1 and yPb = 0 … ((PicHeightInSamplesL + 15) >> 4) − 1,

- The variables xP and yP are set to (xPb << 4) and (yPb << 4), respectively,
- The variables predMode[xP][yP], refIdxLX[xP][yP], mvLX[xP][yP] and refPicOrderCntLX[xP][yP], and predFlagLX[xP][yP], with X = 0,1, of the resampled picture are derived by invoking inter layer motion derivation process specified in

subclause G.8.1.4.2.1 with the luma location (xP, yP), predModeRL, refIdxLXRL, mvLXRL, refPicOrderCntLXRL, and predFlagLXRL, with X = 0,1, given as input.

***~~G.8.1.4.2.1~~ G.8.1.5.1* Derivation process for inter layer motion**

Inputs to this process are

- a luma location (xP, yP) specifying the top-left sample of the current luma prediction block relative to the top-left luma sample of the current picture,
- the reference layer prediction modearray predModeRL,
- the reference layer reference index arrays refIdxL0RL and refIdxL1RL
- the reference layer motion vector arrays mvL0RL and mvL1RL
- the reference layer reference picture order counts arrays refPicOrderCntL0RL and refPicOrderCntL1RL
- the reference layer prediction list utilization flag arrays predFlagL0RL and predFlagL1RL.

Outputs of this process are

- a derived prediction mode predMode,
- two derived motion vectors mvL0 and mvL1
- two derived reference indices refIdxL0 and refIdxL1
- two derived reference picture order counts refPicOrderCntL0 and refPicOrderCntL1
- two derived prediction list utilization flags predFlagL0 and predFlagL1.

The variables predMode, mvLX, refIdxLX, refPicOrderCntLX, and predFlagLX are derived as follows.

1. The center location (xPCtr, yPCtr) of the luma prediction block is derived as follows

$$xPCtr = xP + 8 \quad \text{(G-39)}$$

$$yPCtr = yP + 8 \quad \text{(G-40)}$$

2. The derivation process for reference layer luma sample location specified in subclause G.6.1 is invoked with luma location (xPCtr , yPCtr) given as the inputs and (xRef , yRef) as output.
3. The collocated position (xRL, yRL) is derived as follows

$$xRL = (xRef >> 4) << 4 \quad \text{(G-41)}$$

$$yRL = (yRef >> 4) << 4 \quad \text{(G-42)}$$

4. The reference layer motion vector is derived as follows
   - If (xRL < 0) or (xRL >= RefLayerPicWidthInSamplesL) or (yRL < 0) or (yRL >= RefLayerPicHeightInSamplesL), predMode[xP][yP] is set to MODE_INTRA.
   - Otherwise, predMode[xP][yP] is derived as follows

$$predMode[xP][yP] = predModeRL[xRL][yRL] \quad \text{(G-43)}$$

   - If predMode[xP][yP] is equal to MODE_INTER, for each X = 0, 1, the following applies

$$refIdxLX[xP][yP] = refIdxLXRL[xRL][yRL] \quad \text{(G-44)}$$

refPicOrderCntLX[xP][yP] = refPicOrderCntLXRL[xRL][yRL] (G-45)

predFlagLX[xP][yP] = predFlagLXRL[xRL][yRL] (G-46)

– If ScaledRefLayerPicWidthInSamplesL is not equal to RefLayerPicWidthInSamplesL, mvLX[xP][yP][0] is derived as follows:

scaleFactorMVX = Clip3(−4096, 4095, ((ScaledRefLayerPicWidthInSamplesL << 8) + (RefLayerPicWidthInSamplesL >> 1)) / RefLayerPicWidthInSamplesL) (G-47)

mvLX[xP][yP][0] = Clip3(−32768, 32767, Sign(scaleFactorMVX * mvLXRL[xRL][yRL][0]) * ((Abs (scaleFactorMVX * mvLXRL[xRL][yRL][0]) + 127) >> 8)) (G-48)

– Otherwise,

mvLX[xP][yP][0] = mvLXRL[xRL][yRL][0] (G-49)

– If ScaledRefLayerPicHeightInSamplesL is not equal to RefLayerPicHeightInSamplesL, mvLX[xP][yP][1] is derived as follows:

scaleFactorMVY = Clip3(−4096, 4095, ((ScaledRefLayerPicHeightInSamplesL << 8) + (RefLayerPicHeightInSamplesL >> 1)) / RefLayerPicHeightInSamplesL) (G-50)

mvLX[xP][yP][1] = Clip3(−32768, 32767, Sign(scaleFactorMVY * mvLXRL[xRL][yRL][1]) * ((Abs (scaleFactorMVY * mvLXRL[xRL][yRL][1]) + 127) >> 8)) (G-51)

– Otherwise,

mvLX[xP][yP][1] = mvLXRL[xRL][yRL][1] (G-52)

– Otherwise, if predMode[xP][yP] is equal to MODE_INTRA
  – both components of mvL0[xP][yP] and mvL1[xP][yP] are set to 0,
  refIdxL0[xP][yP] and refIdxL1[xP][yP] are set to –1,
  refPicOrderCntL0[xP][yP] and refPicOrderCntL1[xP][yP] are set to –1,
  predFlagL0[xP][yP] and predFlagL1[xP][yP] are set to 0.

G.11.1.3 Scalable Main profile

Bitstreams conforming to the scalable main profile shall obey the following constraints:

– The picture resampling process as specified in subclauses G.8.1.4. shall not be invoked more than once for decoding of each particular picture.
– *The resampling process of picture motion field as specified in subclause G.8.1.5 shall not be invoked more than once for decoding of each particular picture.*
– When avc_base_layer_flag equal to 1, it is a requirement of bitstream conformance that MotionPredRefLayerId[iNuhLId][mIdx] shall not be equal to 0 for iNuhLId equal to any value of

nuh_layer_id present in the bitstream and any value of mIdx in the range of 0 to NumMotionPredRefLayers[iNuhLId] – 1, inclusive.

Table 2 – Example 2

Bitstream Constraints

[00122] As explained above, the techniques can also provide and/or process bitstream constraints relating to inter-layer prediction types. The term or expression "bitstream constraint" is a broad term and/or expression intended to have its broadest ordinary meaning. In one embodiment, a bitstream constraint can refer to a rule that an encoder or a decoder should follow to be compliant with a certain standard. For example, the conformant to a certain standard bitstream should not contain elements (e.g., syntax elements) that violate the constraint. In the case of a constraint violation, the bitstream is treated as not conformant and may not be decoded by a decoder.

[00123] More specifically, the techniques can provide and/or process a bitstream constraint that the collocated reference index (e.g., collocated_ref_idx) can only refer to an ILRP used for at least ILMP. The techniques can also provide and/or process a bitstream constraint that reference index (e.g., ref_idx) can only refer to an ILRP used for at least ILSP. In some embodiments, the bitstream constraints can be defined as follows. For example,

- **collocated_ref_idx** specifies the reference index of the collocated picture used for temporal motion vector prediction.
  - When slice_type is equal to P or when slice_type is equal to B and collocated_from_l0 is equal to 1, collocated_ref_idx refers to a picture in list 0, and the value of collocated_ref_idx should be in the range of 0 to num_ref_idx_l0_active_minus1, inclusive.
  - When slice_type is equal to B and collocated_from_l0 is equal to 0, collocated_ref_idx refers to a picture in list 1, and the value of collocated_ref_idx should be in the range of 0 to num_ref_idx_l1_active_minus1, inclusive.
- It is a requirement of bitstream conformance that the picture referred to by collocated_ref_idx should be the same for all slices of a coded picture.
- *Let refLayerId be the value of nuh_layer_id of the picture referred to by collocated_ref_idx, and currLayerId be the value of nuh_layer_id of the*

current picture. It is a requirement of bitstream conformance that MotionPredEnabledFlag[currLayerId][refLayerId] should be equal to 1.

- **ref_idx_l0**[x0][y0] specifies the list 0 reference picture index for the current prediction unit. The array indices x0, y0 specify the location (x0, y0) of the top-left luma sample of the considered prediction block relative to the top-left luma sample of the picture.
  - When ref_idx_l0[x0][y0] is not present it is inferred to be equal to 0.
  - *Let refLayerId be the value of nuh_layer_id of the picture referred to by ref_idx_l0[x0][y0], and currLayerId be the value of nuh_layer_id of the current picture. It is a requirement of bitstream conformance that SamplePredEnabledFlag[currLayerId][refLayerId] should be equal to 1.*
  - **ref_idx_l1**[x0][y0] has the same semantics as ref_idx_l0, with l0 and list 0 replaced by l1 and list 1, respectively.

[00124] Certain details relating to the techniques are explained below in reference to FIG. 4 and FIG. 5. Various term used throughout this disclosure are broad terms having their ordinary meaning. In addition, in some embodiments, certain terms relate to the following video concepts. A picture can refer to video picture as that term is used in current standards (e.g., HEVC, SHVC). Methods described with respect to FIG. 4 and FIG. 5 may be implemented by computing hardware. In some embodiments, computing hardware can include one or computing devices comprising computer hardware.

Method for Independent Control of Interlayer Motion Prediction Reference Resampling and Interlayer Sample Prediction Reference Resampling

[00125] FIG. 4 is a flowchart illustrating an example method for independent control of inter-layer motion prediction reference resampling and inter-layer sample prediction reference resampling, according to aspects of this disclosure. The process 400 may be performed by an encoder (e.g., the encoder as shown in FIG. 2A, 2B, etc.), a decoder (e.g., the decoder as shown in FIG. 3A, 3B, etc.), or any other component, depending on the embodiment. The blocks of the process 400 are described with respect to the decoder 31 in FIG. 3B, but the process 400 may be performed by other components, such as an encoder, as mentioned above. The layer 1 video decoder 30B of the decoder 31 and/or the layer 0 decoder 30A of the decoder 31 may perform the

process 400, depending on the embodiment. All embodiments described with respect to FIG. 4 may be implemented separately, or in combination with one another. Certain details relating to the process 400 are explained above.

[00126] The process 400 starts at block 401. The decoder 31 can include a memory (e.g., reference frame memory 82) for storing video information.

[00127] At block 402, the decoder 31 identifies a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising one or more of inter layer motion prediction (ILMP) or inter layer sample prediction (ILSP).

[00128] At block 403, the decoder 31 controls: (1) a number of pictures that may be resampled and used to predict the current picture using ILMP and (2) a number of pictures that may be resampled and used to predict the current picture using ILSP. The number of pictures that may be resampled and used to predict the current picture using ILMP can be controlled independent of the number of pictures that may be resampled and used to predict the current picture using ILSP. For example, the decoder 31 can control the number of pictures that may be resampled and used to predict the current picture using ILMP independent of the number of pictures that may be resampled and used to predict the current picture using ILSP.

[00129] The term or expression "controlled independent," "independent control," or variations thereof is a broad term and/or expression intended to have its broadest ordinary meaning. To facilitate discussion, the term or expression "independent control" will be used in the following description. In one embodiment, independent control can refer to affecting or setting the number of pictures that may be resampled and used to predict the current picture using ILMP without affecting or setting the number of pictures that may be resampled and used to predict the current picture using ILSP, and vice versa.

[00130] In another embodiment, independent control can refer to having a separate limit for the number of pictures that may be resampled and used to predict the current picture using ILMP and the number of pictures that may be resampled and used to predict the current picture using ILSP. The limit on the number of pictures that may be resampled and used to predict the current picture using ILMP and the number of pictures that may be resampled and used to predict the current picture using ILSP can be the same or different, depending on the embodiment. In yet another embodiment,

independent control can refer to not counting a picture that may be resampled (e.g., sample-resampled, motion-resampled, or both) and used to predict the current picture using ILMP towards the limit on the number of pictures that may be resampled and used to predict the current picture using ILP.

[00131] In some embodiments, the number of pictures that may be resampled and used to predict the current picture using ILMP and the number of pictures that may be resampled and used to predict the current picture using ILSP are the same (e.g., both are equal to 1). In other embodiments, the number of pictures that may be resampled and used to predict the current picture using ILMP and the number of pictures that may be resampled and used to predict the current picture using ILSP are different.

[00132] In certain embodiments, the decoder 31 predicts the current picture using at least one resampled picture. The at least one resampled picture can be used to predict the current picture using ILMP, ILSP, or both, depending on the embodiment.

[00133] The process 400 ends at block 404. Blocks may be added and/or omitted in the process 400, depending on the embodiment, and blocks of the process 400 may be performed in different orders, depending on the embodiment. Any features and/or embodiments described with respect to resampling in this disclosure may be implemented separately or in any combination thereof. For example, any features and/or embodiments described in connection with FIG. 4 may be implemented in any combination with any features and/or embodiments described in connection with FIG. 5, and vice versa.

Method for Processing Bitstream Constraints Relating to Inter-Layer Prediction Types

[00134] FIG. 5 is a flowchart illustrating an example method for processing bitstream constraints relating to inter-layer prediction types. The process 500 may be performed by an encoder (e.g., the encoder as shown in FIG. 2A, 2B, etc.), a decoder (e.g., the decoder as shown in FIG. 3A, 3B, etc.), or any other component, depending on the embodiment. The blocks of the process 500 are described with respect to the encoder 21 in FIG. 3B, but the process 500 may be performed by other components, such as a decoder, as mentioned above. The layer 1 video encoder 20B of the encoder 21 and/or the layer 0 encoder 20A of the encoder 21 may perform the process 500, depending on the embodiment. All embodiments described with respect to FIG. 5 may

be implemented separately, or in combination with one another. Certain details relating to the process 500 are explained above, e.g., with respect to FIG. 4.

[00135] The process 500 starts at block 501. The encoder 21 can include a memory (e.g., reference frame memory 82) for storing video information.

[00136] At block 502, the encoder 21 identifies a current picture to be predicted using at least one type of inter layer prediction (ILP). The type of ILP can include inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both.

[00137] At block 503, the encoder 21, when the current picture is to be predicted using at least ILMP, processes a collocated reference index value associated with the current picture, wherein the collocated reference index value indicates a first reference picture that is used in predicting the current picture using ILP. In some embodiments, the collocated reference index value can refer to the value of the collocated reference index. In certain embodiments, the collocated reference index can also be referred to as the collocated reference index value.

[00138] At block 504, the encoder 21 determines whether the first reference picture indicated by the collocated reference index value is enabled for ILMP. For example, the encoder 21 can determine whether the first reference picture is enabled for ILMP when the current picture is to be predicted using at least ILMP. In some embodiments, the encoder 21 determines whether the first reference picture is enabled for ILMP by determining a value of the motion prediction enabled flag for the first reference picture. For example, the encoder 21 can determine that the first reference picture is enabled for ILMP when the motion prediction enabled flag value is equal to 1. In another example, the encoder 21 can determine that the first reference picture is not enabled for ILMP when the motion prediction enabled flag value is equal to 0. In other embodiments, other values of the motion prediction enabled flag value can be used to determine whether the first reference picture is enabled for ILMP or not (e.g., equal to 2, 3, etc.).

[00139] At block 505, the encoder 21, when the current picture is to be predicted using at least ILSP, processes a reference index value associated with a block in the current picture, wherein the reference index value indicates a second reference picture that is used in predicting the block in the current picture using ILP. In some embodiments, the reference index value can refer to the value of the reference index. In

certain embodiments, the reference index can also be referred to as the reference index value.

[00140] At block 506, the encoder 21 determines whether the second reference picture indicated by the reference index value is enabled for ILSP. For example, the encoder 21 can determine whether the second reference picture is enabled for ILSP when the current picture is to be predicted using at least ILSP. In some embodiments, the encoder 21 determines whether the second reference picture is enabled for ILMP by determining a value of the sample prediction enabled flag for the first reference picture. For example, the encoder 21 can determine that the second reference picture is enabled for ILSP when the sample prediction enabled flag value is equal to 1. In another example, the encoder 21 can determine that the second reference picture is not enabled for ILSP when the sample prediction enabled flag value is equal to 0. In other embodiments, other values of the sample prediction enabled flag value can be used to determine whether the second reference picture is enabled for ILSP or not (e.g., equal to 2, 3, etc.).

[00141] In certain embodiments, the encoder 21 signals the collocated reference index value in a bitstream when the first reference picture is enabled for ILMP, or signal the reference index value in the bitstream when the second reference picture is enabled for ILSP, or both. For example, the encoder 21 signals only the collocated reference index value that indicates a reference picture that is enabled for ILMP. Or the encoder 21 signals the reference index value that indicates a reference picture that is enabled for ILSP. Or the encoder 21 can do both. In this manner, only the index values that reference the corresponding type of reference picture (e.g., ILMP enabled for collocated reference index, ILSP enabled for reference index, etc.) may be signaled in the bitstream.

[00142] In some embodiments, the first reference picture and the second reference picture may be the same. For example, a reference picture may be used for both ILMP and ILSP (e.g., have both motion information and samples).

[00143] The process 500 ends at block 507. Blocks may be added and/or omitted in the process 500, depending on the embodiment, and blocks of the process 500 may be performed in different orders, depending on the embodiment. Any features and/or embodiments described with respect to resampling in this disclosure may be implemented separately or in any combination thereof. For example, any features

and/or embodiments described in connection with FIG. 5 may be implemented in any combination with any features and/or embodiments described in connection with FIG. 4, and vice versa.

Inter-Layer Picture Order in Reference Picture List

[00144] In one implementation, three types of inter-layer pictures are possible: motion only inter-layer pictures, sample only inter-layer prediction, and both together. Pictures of all of these types are included into an inter-layer reference picture set. However, pictures with these types may not contribute equally to coding efficiency. For example, pictures used for inter-layer sample prediction can be more important than picture for inter-layer motion prediction only. Accordingly, it may be advantageous to have smaller reference indexes for the pictures for inter-layer sample prediction compared to the pictures for inter-layer motion prediction only.

[00145] In one embodiment, it is suggested to put the pictures for only inter-layer motion prediction at the end of the reference picture set and the initial inter-layer reference picture lists after the pictures for inter-layer sample prediction. So, the order in the reference picture list after all temporal reference pictures and in the inter-layer reference picture set can be as follows by dividing pictures into two sub-sets: pictures for inter-layer sample prediction, pictures for only inter-layer motion prediction. Similarly to the two parts above, alternatively, the order in the reference picture list after all temporal reference pictures and in the inter-layer reference picture set can be as follows by dividing pictures into three sub-sets: pictures for inter-layer sample and motion prediction, pictures for only inter-layer sample prediction, and pictures for only inter-layer motion prediction. Additionally, in each sub-set the ordering can be done in the descending order of the inter-layer picture's nuh_layer_id. Alternatively, the order can follow an explicitly signaled order of the reference layers for inter-layer prediction, which can be signaled in the VPS or elsewhere.

[00146] For the two sub-sets case described above, another type of reference picture sets can be assigned. For example, a sample inter-layer reference picture set can include pictures used for inter-layer sample prediction only or both inter-layer sample prediction and inter-layer motion prediction, and a motion inter-layer reference picture set can include pictures used for inter-layer motion prediction only. Additionally, the ordering can be applied and motion inter-layer reference picture set

can be placed into the initial reference picture lists after sample inter-layer reference picture set. Similarly, for the three sub-set case, the following new inter-layer reference picture set and the ordering can be applied when placing pictures used for inter-layer prediction into the initial reference picture lists: sample and motion inter-layer reference picture set, sample only inter-layer reference picture set, and motion only inter-layer reference picture set. Similarly to the sub-sets, in each new inter-layer reference picture set the picture ordering can be done in the descending order of the inter-layer picture's nuh_layer_id.

Reference Indices Signaling

[00147] The techniques can provide optimization in signaling reference indexes at PU level and co-located reference index at slice level. For example, the total number of reference pictures used in reference index signaling can be adjusted, including only the pictures that can be used for inter prediction, such that the reference index signaled in block (e.g., CU, PU, etc.) level can use fewer bits. In addition, the total number of reference pictures used to define the collocated_ref_idx range can be adjusted to include only the pictures that can be used for TMVP derivation, such that the signaling of collocated_ref_idx can use fewer bits.

[00148] In some embodiments, the variables NumOnlySampleRefIdxLX and NumOnlyMotionRefIdxLX with X being equal to 0 and 1 are derived as follows:

```
NumOnlySampleRefIdxLX = 0
NumOnlyMotionRefIdxLX = 0
for( i = 0; i <= num_ref_idx_lX_active_minus1; i++ ) {
        refLayerId = nuh_layer_id of RefPicListX[ i ]
        if( !SamplePredEnabledFlag[ nuh_layer_id ][ refLayerId ] )
                NumOnlyMotionRefIdxLX ++
        if( !MotionPredEnabledFlag[ nuh_layer_id ][ refLayerId ] )
                NumOnlySampleRefIdxLX ++
}
```

1. PU Reference Signaling

[00149] In one embodiment, **ref_idx_l0**[x0][y0] specifies the list 0 reference picture index for the current prediction unit. The array indices x0, y0 specify the location (x0, y0) of the top-left luma sample of the considered prediction block relative to the top-left luma sample of the picture. **ref_idx_l1**[x0][y0] has the same

semantics as ref_idx_l0, with l0 and list 0 replaced by l1 and list 1, respectively. In certain embodiments, the coding process may be changed from the early versions of SHVC as follows (changes indicated in bold and italics):

prediction_unit(x0, y0, nPbW, nPbH) {	**Descriptor**
…	
if(inter_pred_idc[x0][y0] != PRED_L1) {	
if(num_ref_idx_l0_active_minus1 – *NumOnlyMotionRefIdxL0* > 0)	
ref_idx_l0[x0][y0]	ae(v)
mvd_coding(x0, y0, 0)	
mvp_l0_flag[x0][y0]	ae(v)
}	
if(inter_pred_idc[x0][y0] != PRED_L0) {	
if(num_ref_idx_l1_active_minus1 – *NumOnlyMotionRefIdxL1* > 0)	
ref_idx_l1[x0][y0]	ae(v)
if(mvd_l1_zero_flag && inter_pred_idc[x0][y0] = = PRED_BI) {	
MvdL1[x0][y0][0] = 0	
MvdL1[x0][y0][1] = 0	
} else	
mvd_coding(x0, y0, 1)	
mvp_l1_flag[x0][y0]	ae(v)
}	
…	

[00150] *ref_idx_lX[x0][y0] is adjusted as follows with X being equal to 0 and 1:*

RefIdxNum = ref_idx_lX[x0][y0]
for(i = 0; i <= RefIdxNum; i++) {
refLayerId = nuh_layer_id of the RefPicListX[i]
if(!SamplePredEnabledFlag[nuh_layer_id][refLayerId])
ref_idx_lX[x0][y0]++
}

2. Collocated Reference Index Signaling

[00151] In one embodiment, **collocated_ref_idx** specifies the reference index of the collocated picture used for temporal motion vector prediction.

[00152] When slice_type is equal to P or when slice_type is equal to B and collocated_from_l0 is equal to 1, collocated_ref_idx refers to a picture in list 0, and the value of collocated_ref_idx should be in the range of 0 to num_ref_idx_l0_active_minus1 – *NumOnlySampleRefIdxL0*, inclusive.

[00153] When slice_type is equal to B and collocated_from_l0 is equal to 0, collocated_ref_idx refers to a picture in list 1, and the value of collocated_ref_idx should be in the range of 0 to num_ref_idx_l1_active_minus1 – *NumOnlySampleRefIdxL1*, inclusive.

[00154] It is a requirement of bitstream conformance that the picture referred to by collocated_ref_idx should be the same for all slices of a coded picture.

[00155] *collocated_ref_idx is adjusted as follows:*

```
RefIdxNum = collocated_ref_idx
for( i = 0; i <= RefIdxNum; i++ ) {
        refLayerId = nuh_layer_id of the RefPicListX[ i ]
        if( !MotionPredEnabledFlag[ nuh_layer_id ][ refLayerId ] )
        collocated_ref_idx++
}
```

with X being equal to collocated_from_l0.

Terminology

[00156] While the above disclosure has described particular embodiments, many variations are possible. For example, as mentioned above, the above techniques may be applied to 3D video encoding. In some embodiments of 3D video, a reference layer (e.g., a base layer) includes video information sufficient to display a first view of a video and the enhancement layer includes additional video information relative to the reference layer such that the reference layer and the enhancement layer together include video information sufficient to display a second view of the video. These two views can used to generate a stereoscopic image. As discussed above, motion information from the reference layer can be used to identify additional implicit hypothesis when encoding or decoding a video unit in the enhancement layer, in accordance with aspects of the disclosure. This can provide greater coding efficiency for a 3D video bitstream.

[00157] It is to be recognized that depending on the example, certain acts or events of any of the techniques described herein can be performed in a different sequence, may be added, merged, or left out altogether (e.g., not all described acts or

events are necessary for the practice of the techniques). Moreover, in certain examples, acts or events may be performed concurrently, e.g., through multi-threaded processing, interrupt processing, or multiple processors, rather than sequentially.

[00158] Information and signals disclosed herein may be represented using any of a variety of different technologies and techniques. For example, data, instructions, commands, information, signals, bits, symbols, and chips that may be referenced throughout the above description may be represented by voltages, currents, electromagnetic waves, magnetic fields or particles, optical fields or particles, or any combination thereof.

[00159] The various illustrative logical blocks, modules, circuits, and algorithm steps described in connection with the embodiments disclosed herein may be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, modules, circuits, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans may implement the described functionality in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the present invention.

[00160] The techniques described herein may be implemented in hardware, software, firmware, or any combination thereof. Such techniques may be implemented in any of a variety of devices such as general purposes computers, wireless communication device handsets, or integrated circuit devices having multiple uses including application in wireless communication device handsets and other devices. Any features described as modules or components may be implemented together in an integrated logic device or separately as discrete but interoperable logic devices. If implemented in software, the techniques may be realized at least in part by a computer-readable data storage medium comprising program code including instructions that, when executed, performs one or more of the methods described above. The computer-readable data storage medium may form part of a computer program product, which may include packaging materials. The computer-readable medium may comprise memory or data storage media, such as random access memory (RAM) such as synchronous dynamic random access memory (SDRAM), read-only memory (ROM),

non-volatile random access memory (NVRAM), electrically erasable programmable read-only memory (EEPROM), FLASH memory, magnetic or optical data storage media, and the like. The techniques additionally, or alternatively, may be realized at least in part by a computer-readable communication medium that carries or communicates program code in the form of instructions or data structures and that can be accessed, read, and/or executed by a computer, such as propagated signals or waves.

[00161] The program code may be executed by a processor, which may include one or more processors, such as one or more digital signal processors (DSPs), general purpose microprocessors, an application specific integrated circuits (ASICs), field programmable logic arrays (FPGAs), or other equivalent integrated or discrete logic circuitry. Such a processor may be configured to perform any of the techniques described in this disclosure. A general purpose processor may be a microprocessor; but in the alternative, the processor may be any conventional processor, controller, microcontroller, or state machine. A processor may also be implemented as a combination of computing devices, e.g., a combination of a DSP and a microprocessor, a plurality of microprocessors, one or more microprocessors in conjunction with a DSP core, or any other such configuration. Accordingly, the term "processor," as used herein may refer to any of the foregoing structure, any combination of the foregoing structure, or any other structure or apparatus suitable for implementation of the techniques described herein. In addition, in some aspects, the functionality described herein may be provided within dedicated software modules or hardware modules configured for encoding and decoding, or incorporated in a combined video encoder-decoder (CODEC).

[00162] The coding techniques discussed herein may be embodiment in an example video encoding and decoding system. A system includes a source device that provides encoded video data to be decoded at a later time by a destination device. In particular, the source device provides the video data to destination device via a computer-readable medium. The source device and the destination device may comprise any of a wide range of devices, including desktop computers, notebook (i.e., laptop) computers, tablet computers, set-top boxes, telephone handsets such as so-called “smart” phones, so-called “smart” pads, televisions, cameras, display devices, digital media players, video gaming consoles, video streaming device, or the like. In some

cases, the source device and the destination device may be equipped for wireless communication.

[00163] The destination device may receive the encoded video data to be decoded via the computer-readable medium. The computer-readable medium may comprise any type of medium or device capable of moving the encoded video data from source device to destination device. In one example, computer-readable medium may comprise a communication medium to enable source device 12 to transmit encoded video data directly to destination device in real-time. The encoded video data may be modulated according to a communication standard, such as a wireless communication protocol, and transmitted to destination device. The communication medium may comprise any wireless or wired communication medium, such as a radio frequency (RF) spectrum or one or more physical transmission lines. The communication medium may form part of a packet-based network, such as a local area network, a wide-area network, or a global network such as the Internet. The communication medium may include routers, switches, base stations, or any other equipment that may be useful to facilitate communication from source device to destination device.

[00164] In some examples, encoded data may be output from output interface to a storage device. Similarly, encoded data may be accessed from the storage device by input interface. The storage device may include any of a variety of distributed or locally accessed data storage media such as a hard drive, Blu-ray discs, DVDs, CD-ROMs, flash memory, volatile or non-volatile memory, or any other suitable digital storage media for storing encoded video data. In a further example, the storage device may correspond to a file server or another intermediate storage device that may store the encoded video generated by source device. Destination device may access stored video data from the storage device via streaming or download. The file server may be any type of server capable of storing encoded video data and transmitting that encoded video data to the destination device. Example file servers include a web server (e.g., for a website), an FTP server, network attached storage (NAS) devices, or a local disk drive. Destination device may access the encoded video data through any standard data connection, including an Internet connection. This may include a wireless channel (e.g., a Wi-Fi connection), a wired connection (e.g., DSL, cable modem, etc.), or a combination of both that is suitable for accessing encoded video data stored on a file

server. The transmission of encoded video data from the storage device may be a streaming transmission, a download transmission, or a combination thereof.

[00165] The techniques of this disclosure are not necessarily limited to wireless applications or settings. The techniques may be applied to video coding in support of any of a variety of multimedia applications, such as over-the-air television broadcasts, cable television transmissions, satellite television transmissions, Internet streaming video transmissions, such as dynamic adaptive streaming over HTTP (DASH), digital video that is encoded onto a data storage medium, decoding of digital video stored on a data storage medium, or other applications. In some examples, system may be configured to support one-way or two-way video transmission to support applications such as video streaming, video playback, video broadcasting, and/or video telephony.

[00166] In one example the source device includes a video source, a video encoder, and a output interface. The destination device may include include an input interface, a video decoder, and a display device. The video encoder of source device may be configured to apply the techniques disclosed herein. In other examples, a source device and a destination device may include other components or arrangements. For example, the source device may receive video data from an external video source, such as an external camera. Likewise, the destination device may interface with an external display device, rather than including an integrated display device.

[00167] The example system above merely one example. Techniques for processing video data in parallel may be performed by any digital video encoding and/or decoding device. Although generally the techniques of this disclosure are performed by a video encoding device, the techniques may also be performed by a video encoder/decoder, typically referred to as a "CODEC." Moreover, the techniques of this disclosure may also be performed by a video preprocessor. Source device and destination device are merely examples of such coding devices in which source device generates coded video data for transmission to destination device. In some examples, the source and destination devices may operate in a substantially symmetrical manner such that each of the devices includes video encoding and decoding components. Hence, example systems may support one-way or two-way video transmission between video devices, e.g., for video streaming, video playback, video broadcasting, or video telephony.

[00168] The video source may include a video capture device, such as a video camera, a video archive containing previously captured video, and/or a video feed interface to receive video from a video content provider. As a further alternative, the video source may generate computer graphics-based data as the source video, or a combination of live video, archived video, and computer-generated video. In some cases, if video source is a video camera, source device and destination device may form so-called camera phones or video phones. As mentioned above, however, the techniques described in this disclosure may be applicable to video coding in general, and may be applied to wireless and/or wired applications. In each case, the captured, pre-captured, or computer-generated video may be encoded by the video encoder. The encoded video information may then be output by output interface onto the computer-readable medium.

[00169] As noted the computer-readable medium may include transient media, such as a wireless broadcast or wired network transmission, or storage media (that is, non-transitory storage media), such as a hard disk, flash drive, compact disc, digital video disc, Blu-ray disc, or other computer-readable media. In some examples, a network server (not shown) may receive encoded video data from the source device and provide the encoded video data to the destination device, e.g., via network transmission. Similarly, a computing device of a medium production facility, such as a disc stamping facility, may receive encoded video data from the source device and produce a disc containing the encoded video data. Therefore, the computer-readable medium may be understood to include one or more computer-readable media of various forms, in various examples.

[00170] The input interface of the destination device receives information from the computer-readable medium. The information of the computer-readable medium may include syntax information defined by the video encoder, which is also used by the video decoder, that includes syntax elements that describe characteristics and/or processing of blocks and other coded units, e.g., group of pictures (GOP). A display device displays the decoded video data to a user, and may comprise any of a variety of display devices such as a cathode ray tube (CRT), a liquid crystal display (LCD), a plasma display, an organic light emitting diode (OLED) display, or another type of display device. Various embodiments of the invention have been described. These and other embodiments are within the scope of the following claims.

[00171] Various embodiments of the invention have been described. These and other embodiments are within the scope of the following claims.

55479-16

CLAIMS:

1. An apparatus configured to code video information, the apparatus comprising:

a memory configured to store video data; and

computing hardware operationally coupled to the memory and configured to:

identify a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both;

in response to a determination that the current picture is to be predicted using at least ILMP:

process a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP;

determine whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; and

in response to a determination that the current picture is to be predicted using at least ILSP:

process a reference index value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP; and

determine whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

2. The apparatus of Claim 1, wherein the computing hardware is further configured to determine that the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP.

3. The apparatus of Claim 1, wherein the computing hardware is further configured to determine that the first reference picture is enabled for ILMP when the motion prediction enabled flag value is equal to 1.

4. The apparatus of Claim 1, wherein the computing hardware is further configured to determine that the first reference picture is not enabled for ILMP when the motion prediction enabled flag value is equal to 0.

5. The apparatus of Claim 1, wherein the computing hardware is configured to determine the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP.

6. The apparatus of Claim 5, wherein the computing hardware is further configured to determine that the second reference picture is enabled for ILSP when the sample prediction enabled flag value is equal to 1.

7. The apparatus of Claim 5, wherein the computing hardware is further configured to determine that the second reference picture is not enabled for ILSP when the sample prediction enabled flag value is equal to 0.

8. The apparatus of Claim 1, wherein the computing hardware is further configured to signal the collocated reference index value in a bitstream when the first reference picture is enabled for ILMP, or signal the reference index value in the bitstream when the second reference picture is enabled for ILSP, or both.

9. The apparatus of Claim 1, wherein the computing hardware is further configured to receive the collocated reference index or the reference index in a bitstream.

10. The apparatus of Claim 1, wherein the computing hardware is further configured to encode the collocated reference index value in a bitstream when the first

reference picture is enabled for ILMP, or encode the reference index value in the bitstream when the second reference picture is enabled for ILSP, or both.

11. The apparatus of Claim 1, wherein the computing hardware is further configured to decode the collocated reference index or the reference index in a bitstream.

12. The apparatus of Claim 1, wherein the apparatus is selected from a group consisting of one or more of: a desktop computer, a notebook computer, a laptop computer, a tablet computer, a set-top box, a telephone handset, a smart phone, a smart pad, a television, a camera, a display device, a digital media player, a video gaming console, and a video streaming device.

13. A method of coding video information, the method comprising:

identifying a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both; and

performing at least one of:

(i) in response to a determination that the current picture is to be predicted using at least ILMP:

processing a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP; and

determining whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; or

(ii) in response to a determination that the current picture is to be predicted using at least ILSP:

processing a reference value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP; and

determining whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

14. The method of Claim 13, further comprising determining that the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP.

15. The method of Claim 14, further comprising determining that the first reference picture is enabled for ILMP when the motion prediction enabled flag value is equal to 1.

16. The method of Claim 14, further comprising determining that the first reference picture is not enabled for ILMP when the motion prediction enabled flag value is equal to 0.

17. The method of Claim 13, further comprising determining that the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP.

18. The method of Claim 17, further comprising determining that the second reference picture is enabled for ILSP when the sample prediction enabled flag value is equal to 1.

19. The method of Claim 17, further comprising determining that the second reference picture is not enabled for ILSP when the sample prediction enabled flag value is equal to 0.

20. The method of Claim 13, further comprising signaling the collocated reference index value in a bitstream when the first reference picture is enabled for ILMP, or signaling

the reference index value in the bitstream when the second reference picture is enabled for ILSP, or both.

21. The method of Claim 13, further comprising receiving the collocated reference index or the reference index in a bitstream.

22. The method of Claim 13, further comprising encoding the collocated reference index value in a bitstream when the first reference picture is enabled for ILMP, or encoding the reference index value in the bitstream when the second reference picture is enabled for ILSP, or both.

23. The method of Claim 13, further comprising decoding the collocated reference index or the reference index in a bitstream.

24. A non-transitory computer readable medium comprising instructions that when executed on a processor comprising computer hardware cause the processor to:

identify a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both; and

perform at least one of:

(i) in response to a determination that the current picture is to be predicted using at least ILMP:

process a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP; and

determine whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; or

(ii) in response to a determination that the current picture is to be predicted using at least ILSP:

process a reference index value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP; and

determine whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

25. The computer readable medium of Claim 24, further comprising instructions that cause the processor to determine that the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP.

26. The computer readable medium of Claim 24, further comprising instructions that cause the processor to determine that the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP.

27. An apparatus configured to code video information, the apparatus comprising:

means for storing video data;

means for identifying a current picture to be predicted using at least one type of inter layer prediction (ILP), the type of ILP comprising inter layer motion prediction (ILMP), or inter layer sample prediction (ILSP), or both;

means for, in response to a determination that the current picture is to be predicted using at least ILMP, processing a collocated reference index value indicative of an index of a collocated picture of the current picture used for temporal motion vector prediction, the collocated reference index value associated with the current picture, wherein the collocated reference index value specifies a first reference picture that is to be used in predicting the current picture using ILP,

the means for processing the collocated reference index value further configured to determine whether the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP based on a motion prediction enabled flag associated with the first reference picture; and

means for, in response to a determination that the current picture is to be predicted using at least ILSP, processing a reference index value indicative of an index of a reference picture for a block in the current picture, the reference index value associated with the block in the current picture, wherein the reference index value specifies a second reference picture that is to be used in predicting the block in the current picture using ILP,

the means for processing the reference index value further configured to determine whether the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP based on a sample prediction enabled flag associated with the second reference picture.

28. The apparatus of Claim 27, wherein the means for determining whether the first reference picture is enabled for ILMP is configured to determine that the first reference picture specified by the collocated reference index value is allowed to be used for predicting the current picture using ILMP.

29. The apparatus of Claim 27, wherein the means for determining whether the second reference picture is enabled for ILSP is configured to determine that the second reference picture specified by the reference index value is allowed to be used for predicting the block in the current picture using ILSP.

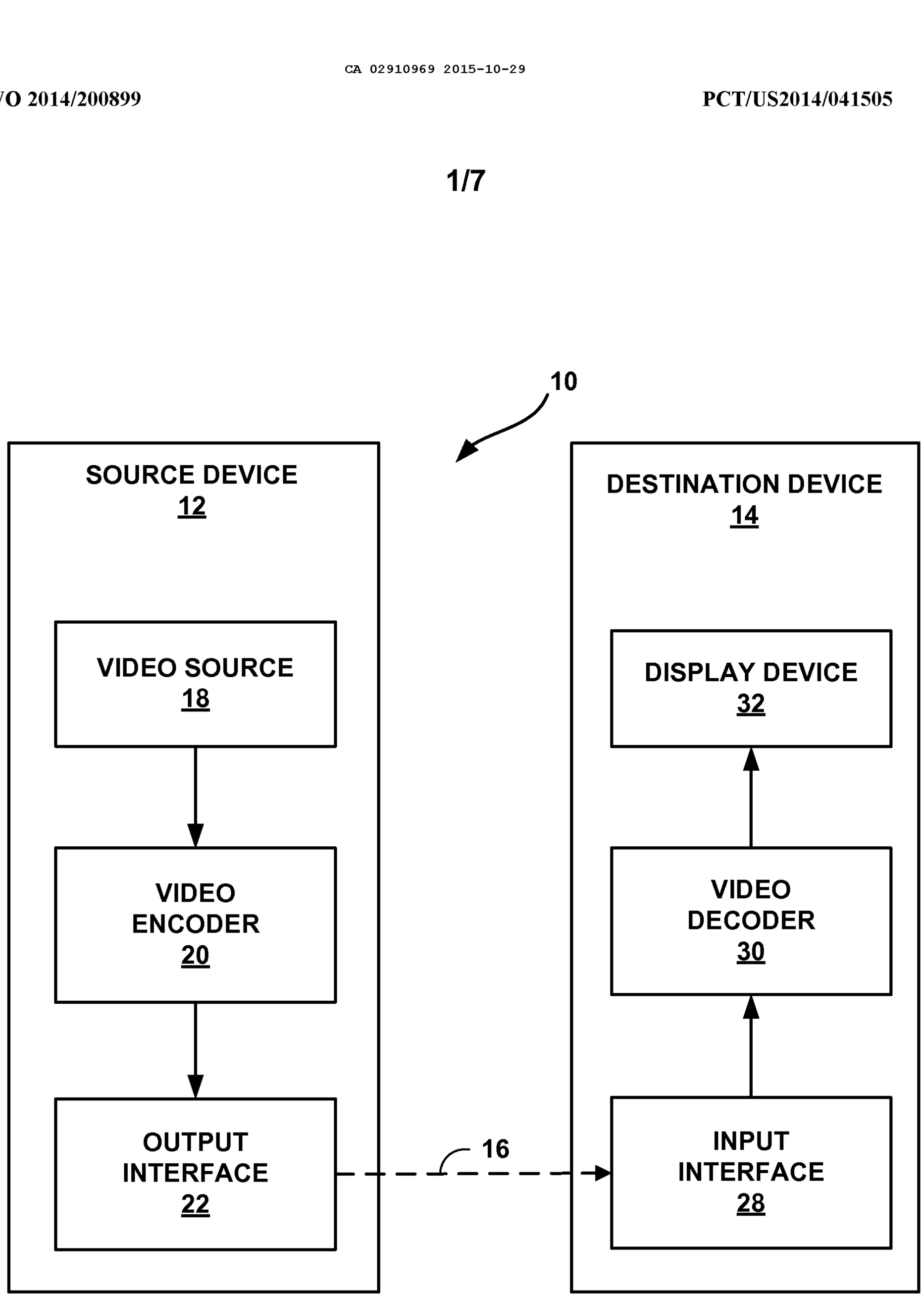

10
SOURCE DEVICE
12
VIDEO SOURCE
18
VIDEO
ENCODER
20
OUTPUT
INTERFACE
22
16
DESTINATION DEVICE
14
DISPLAY DEVICE
32
VIDEO
DECODER
30
INPUT
INTERFACE
28

FIG. 1

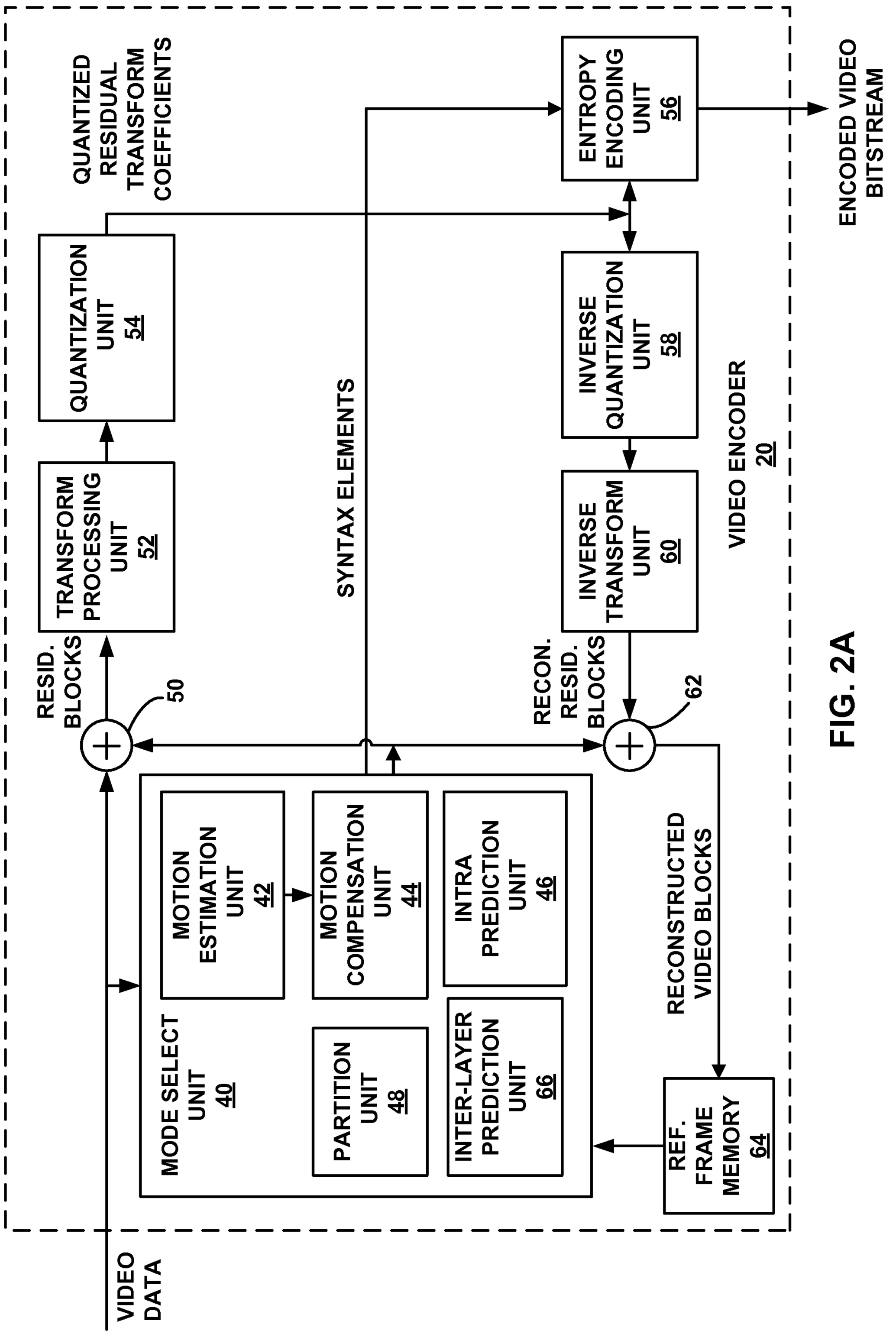

VIDEO DATA
RESID. BLOCKS
50
TRANSFORM PROCESSING UNIT 52
QUANTIZATION UNIT 54
QUANTIZED RESIDUAL TRANSFORM COEFFICIENTS
SYNTAX ELEMENTS
ENTROPY ENCODING UNIT 56
ENCODED VIDEO BITSTREAM
INVERSE QUANTIZATION UNIT 58
INVERSE TRANSFORM UNIT 60
RECON. RESID. BLOCKS
62
RECONSTRUCTED VIDEO BLOCKS
MODE SELECT UNIT 40
MOTION ESTIMATION UNIT 42
MOTION COMPENSATION UNIT 44
PARTITION UNIT 48
INTRA PREDICTION UNIT 46
INTER-LAYER PREDICTION UNIT 66
REF. FRAME MEMORY 64
VIDEO ENCODER 20

FIG. 2A

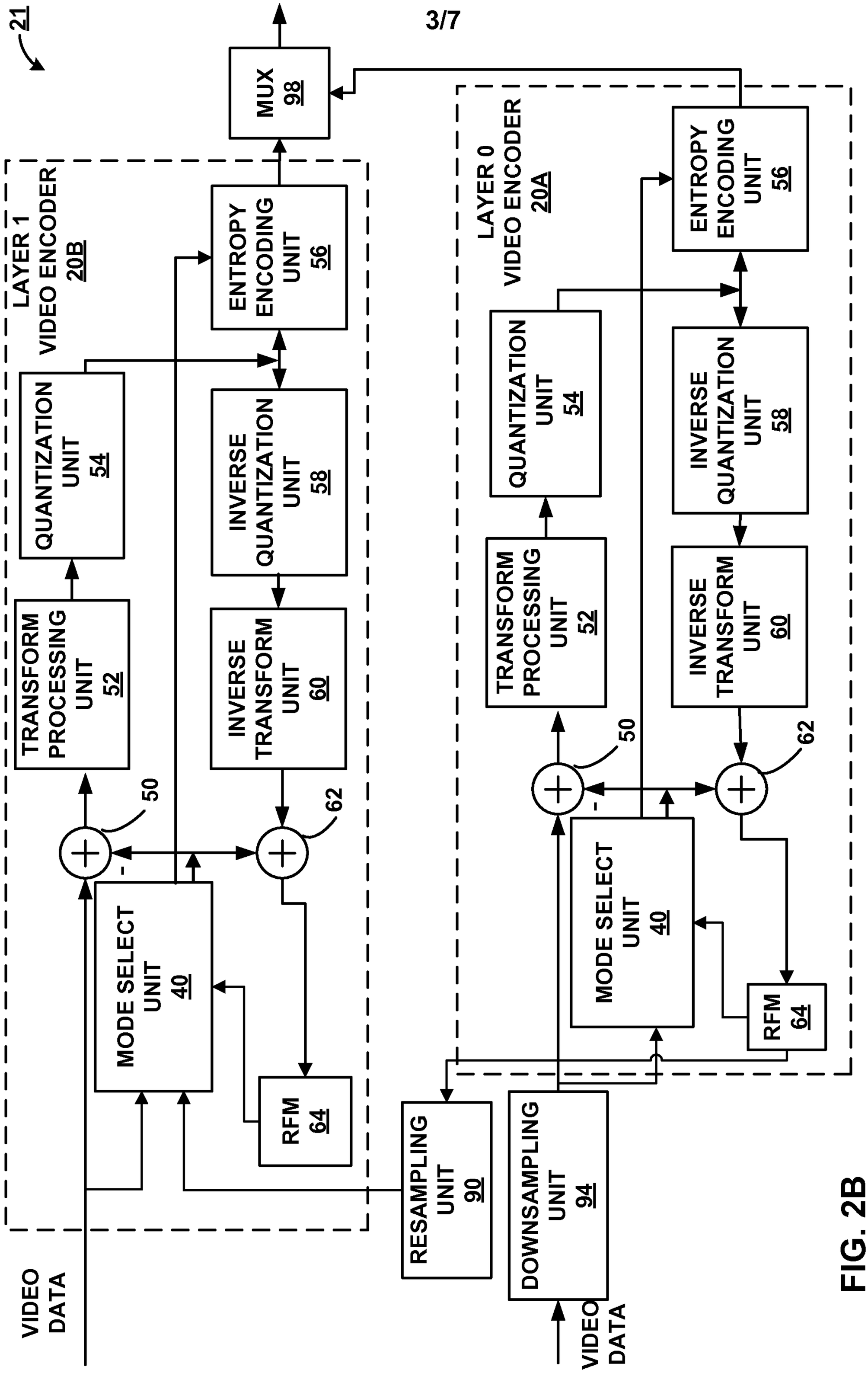

21
MUX
98
LAYER 1
VIDEO ENCODER
20B
LAYER 0
VIDEO ENCODER
20A
ENTROPY
ENCODING
UNIT
56
ENTROPY
ENCODING
UNIT
56
QUANTIZATION
UNIT
54
QUANTIZATION
UNIT
54
INVERSE
QUANTIZATION
UNIT
58
INVERSE
QUANTIZATION
UNIT
58
TRANSFORM
PROCESSING
UNIT
52
TRANSFORM
PROCESSING
UNIT
52
INVERSE
TRANSFORM
UNIT
60
INVERSE
TRANSFORM
UNIT
60
50
50
62
62
MODE SELECT
UNIT
40
MODE SELECT
UNIT
40
RFM
64
RFM
64
RESAMPLING
UNIT
90
DOWNSAMPLING
UNIT
94
VIDEO
DATA
VIDEO
DATA

FIG. 2B

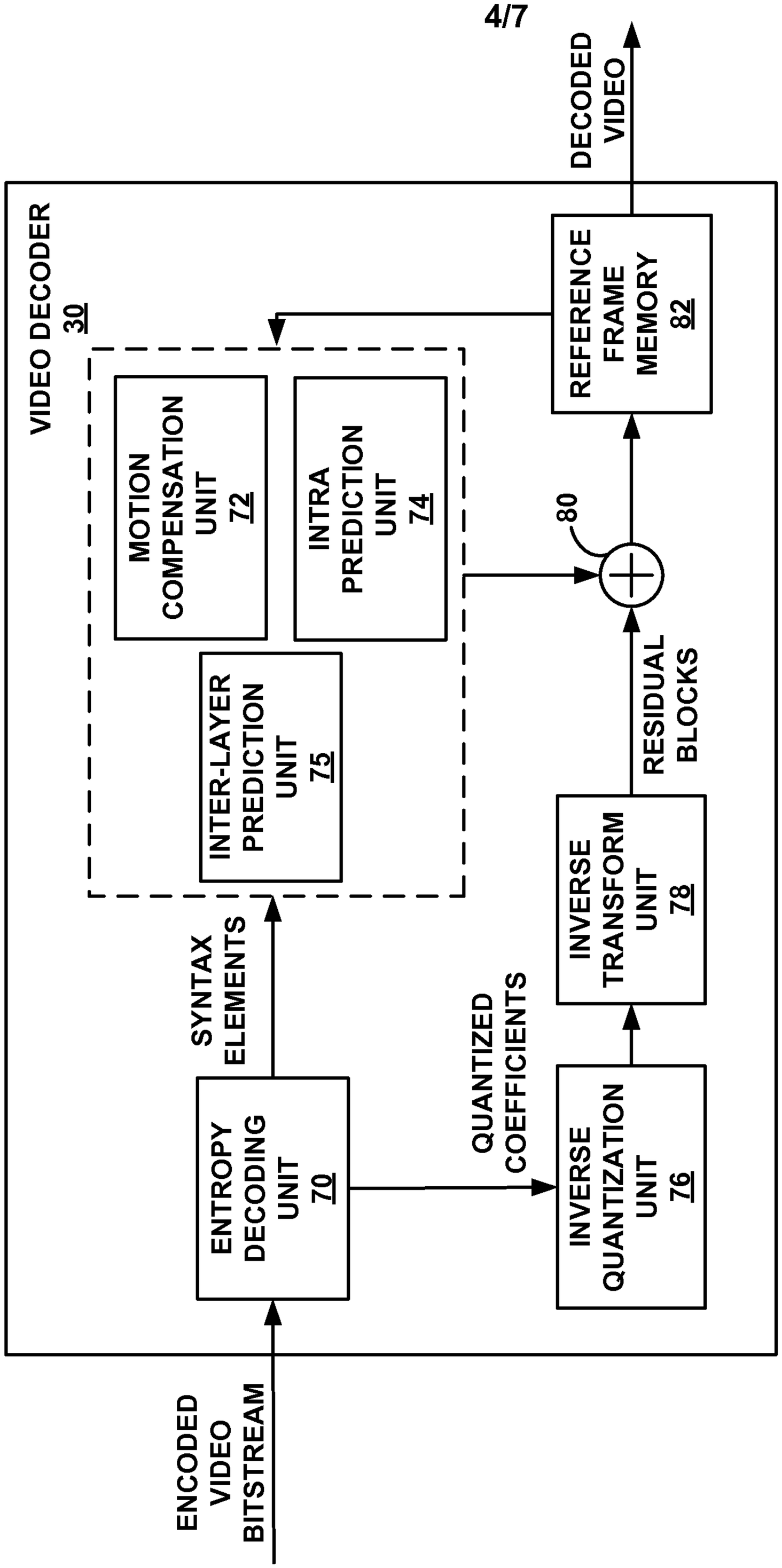

ENCODED VIDEO BITSTREAM
VIDEO DECODER 30
ENTROPY DECODING UNIT 70
SYNTAX ELEMENTS
MOTION COMPENSATION UNIT 72
INTRA PREDICTION UNIT 74
INTER-LAYER PREDICTION UNIT 75
INVERSE QUANTIZATION UNIT 76
QUANTIZED COEFFICIENTS
INVERSE TRANSFORM UNIT 78
RESIDUAL BLOCKS
80
REFERENCE FRAME MEMORY 82
DECODED VIDEO

FIG. 3A

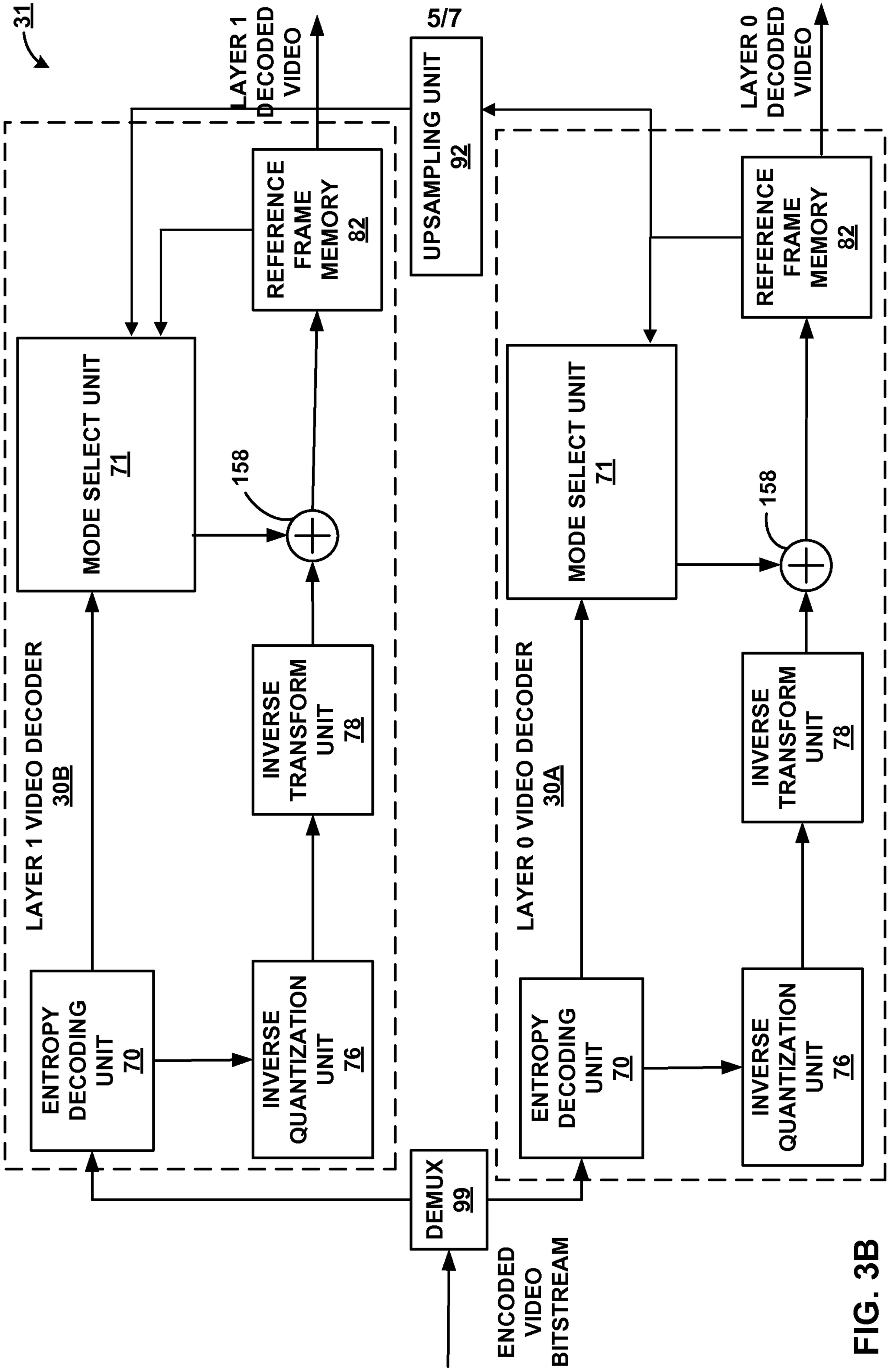

31
LAYER 1 DECODED VIDEO
5/7
UPSAMPLING UNIT 92
LAYER 0 DECODED VIDEO
REFERENCE FRAME MEMORY 82
MODE SELECT UNIT 71
158
INVERSE TRANSFORM UNIT 78
LAYER 1 VIDEO DECODER 30B
LAYER 0 VIDEO DECODER 30A
ENTROPY DECODING UNIT 70
INVERSE QUANTIZATION UNIT 76
DEMUX 99
ENCODED VIDEO BITSTREAM

FIG. 3B

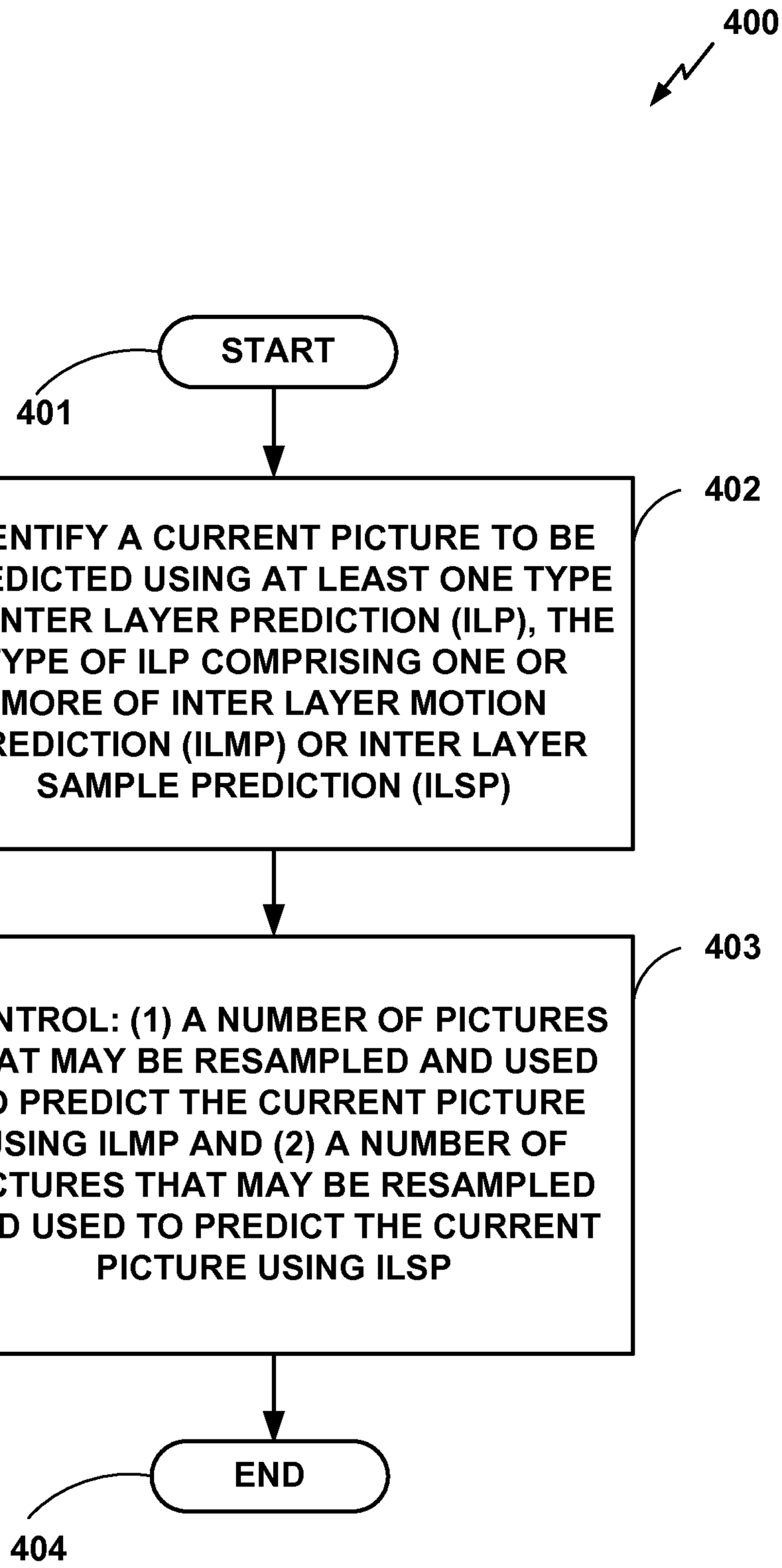

400
START
401
IDENTIFY A CURRENT PICTURE TO BE PREDICTED USING AT LEAST ONE TYPE OF INTER LAYER PREDICTION (ILP), THE TYPE OF ILP COMPRISING ONE OR MORE OF INTER LAYER MOTION PREDICTION (ILMP) OR INTER LAYER SAMPLE PREDICTION (ILSP)
402
CONTROL: (1) A NUMBER OF PICTURES THAT MAY BE RESAMPLED AND USED TO PREDICT THE CURRENT PICTURE USING ILMP AND (2) A NUMBER OF PICTURES THAT MAY BE RESAMPLED AND USED TO PREDICT THE CURRENT PICTURE USING ILSP
403
END
404

FIG. 4

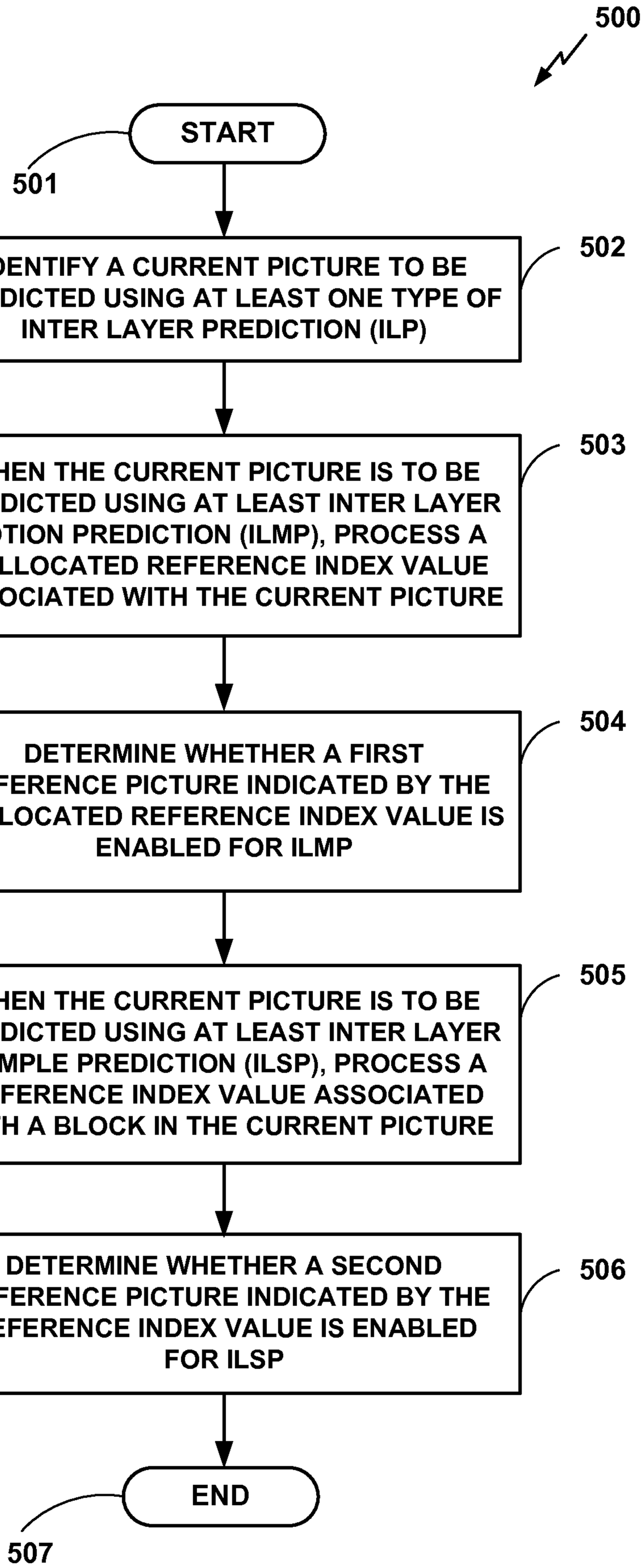
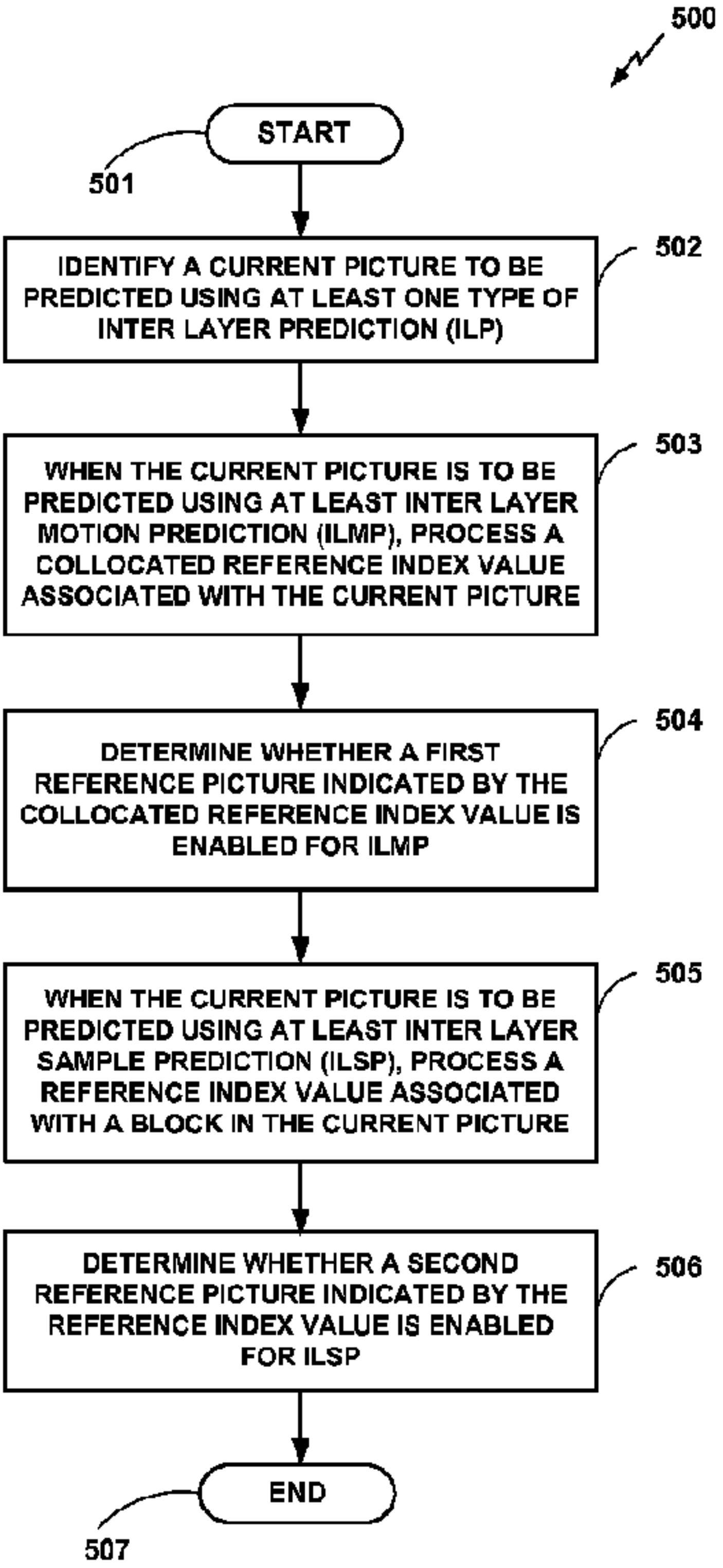

500
START
501
IDENTIFY A CURRENT PICTURE TO BE PREDICTED USING AT LEAST ONE TYPE OF INTER LAYER PREDICTION (ILP)
502
WHEN THE CURRENT PICTURE IS TO BE PREDICTED USING AT LEAST INTER LAYER MOTION PREDICTION (ILMP), PROCESS A COLLOCATED REFERENCE INDEX VALUE ASSOCIATED WITH THE CURRENT PICTURE
503
DETERMINE WHETHER A FIRST REFERENCE PICTURE INDICATED BY THE COLLOCATED REFERENCE INDEX VALUE IS ENABLED FOR ILMP
504
WHEN THE CURRENT PICTURE IS TO BE PREDICTED USING AT LEAST INTER LAYER SAMPLE PREDICTION (ILSP), PROCESS A REFERENCE INDEX VALUE ASSOCIATED WITH A BLOCK IN THE CURRENT PICTURE
505
DETERMINE WHETHER A SECOND REFERENCE PICTURE INDICATED BY THE REFERENCE INDEX VALUE IS ENABLED FOR ILSP
506
END
507

FIG. 5


500
START
501
IDENTIFY A CURRENT PICTURE TO BE PREDICTED USING AT LEAST ONE TYPE OF INTER LAYER PREDICTION (ILP)
502
WHEN THE CURRENT PICTURE IS TO BE PREDICTED USING AT LEAST INTER LAYER MOTION PREDICTION (ILMP), PROCESS A COLLOCATED REFERENCE INDEX VALUE ASSOCIATED WITH THE CURRENT PICTURE
503
DETERMINE WHETHER A FIRST REFERENCE PICTURE INDICATED BY THE COLLOCATED REFERENCE INDEX VALUE IS ENABLED FOR ILMP
504
WHEN THE CURRENT PICTURE IS TO BE PREDICTED USING AT LEAST INTER LAYER SAMPLE PREDICTION (ILSP), PROCESS A REFERENCE INDEX VALUE ASSOCIATED WITH A BLOCK IN THE CURRENT PICTURE
505
DETERMINE WHETHER A SECOND REFERENCE PICTURE INDICATED BY THE REFERENCE INDEX VALUE IS ENABLED FOR ILSP
506
END
507